GENOMIC LANDSCAPES OF CARDIAC ARRHYTHMIA AND HEART REGENERATION

A Dissertation

by

MIN ZHANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Fen Wang
Committee Members,	James F. Martin
	Jiang Chang
	Thomas A. Cooper
Head of Department,	Van G. Wilson

August 2015

Major Subject: Medical Sciences

ABSTRACT

Use of high-throughput sequencing technology has revolutionized our ability to study heart development and regeneration. However, extracting crucial information from large-scale datasets is challenging. Here we demonstrated using bioinformatic analysis to identify genetic features causing human cardiac arrhythmias, as well as key components of heart regeneration.

Pitx2 is the homeobox gene located in proximity to the human 4q25 familial atrial fibrillation locus. Previous work focused on *Pitx2* developmental functions that predispose to atrial fibrillation. It is unknown whether *Pitx2* has distinct postnatal functions. Furthermore, it is unclear whether the 4q25 locus directly controls the expression of *Pitx2* via physical interaction that forms three-dimensional chromatin structure. Inactivating *Pitx2* in the postnatal atrium, unstressed adult *Pitx2* homozygous mutant mice display sinus node dysfunction, an atrial fibrillation risk factor in human patients. An integrated genomics approach revealed Pitx2 target genes encoding cell junction proteins, ion channels, and critical transcriptional regulators. Importantly, many Pitx2 target genes have been implicated in human atrial fibrillation by genome wide association studies. Moreover, chromatin conformation capture sequencing (4C-Seq) demonstrated a direct interaction of *Pitx2* promoters and atrial-fibrillation-associated region.

The mammalian heart regenerates poorly, and damage commonly leads to heart failure. Hippo signaling is kinase cascade that regulates organ size during development and prevents adult mammalian cardiomyocyte regeneration by inhibiting the transcription factor Yap. To identify Yap target genes that are activated during cardiomyocyte renewal and regeneration, we performed Yap chromatin immunoprecipita-

tion sequencing (ChIP-Seq) and mRNA expression profiling in Hippo-deficient mouse hearts. We found that Yap directly regulated genes encoding cell cycle progression proteins, as well genes encoding proteins that promote F-actin polymerization and link the actin cytoskeleton to the extracellular matrix. Border-zone cardiomyocytes of injured Hippo-deficient mouse hearts showed cellular protrusions, indicating cytoskeletal remodeling.

Our findings, revealing an independent postnatal role of *Pitx2* in arrhythmias, unveil direct Pitx2 target genes that include channel and calcium handling genes as well as genes that stabilize the intercalated disc in postnatal atrium. In the context of mammalian cardiomyocyte regeneration, in addition to activating cell cycle genes, Yap regulated genes enhance cytoskeletal remodeling and the cellular response to local mechanical microenvironment.

DEDICATION

To my parents

ACKNOWLEDGEMENTS

I would not have been able to accomplish the academic endeavor without the guidance and support from numerous people. Firstly, I want to thank my advisor, Dr. James Martin. Joining Dr. Martin's lab is one of the best decisions I have ever made. He gives me freedom and great encouragement to pursue my research interest in bioinformatics. His dedication to work and vision in science has set the standard for the rest of my career.

I am extremely grateful to the members of my committee, Drs Fen Wang, Jiang Chang and Thomas Cooper. They have been kind, insightful, and generous with their time during the past five years.

I would like to thank Chinese Scholarship Council and East China Normal University for offering me the extraordinary opportunity to study abroad and providing generous financial support.

Thanks to Texas A&M Institute of Biosciences and Technology department and Baylor College of Medicine Molecular Physiology department for creating the outstanding research environments. Especially, Cindy Lewis, Janis Bender, Samantha Del Castillo from IBT; Debbie Colbert, Lynda Attaway, Janet Tran, and Ross Adkins from Baylor for their great support over the years.

I want to thank all the past and current lab members. Todd Heallen, who not only patiently taught me everything when I first joined the lab, but also introduced me to the American cultures and all its nuances. Jun Wang, Margarita Bonilla, Ela Klysik, and Tien Tran, all of whom treat me like family and happily helped me anytime I asked. Paul Swinton, who performed all my sequencing and micro-injection experiments and someone whom I can always share a good laugh with. Ye Tao, Yuka Morikawa, Todd

Heallen, John Leach, Yan Bai, Yang Xiao, Peter Kahr, Lele Li, and Ge Tao, all of whom I collaborated with and contributed to a lot of work I present in this dissertation. I also want to thank all of our collaborators for working on great science together, including Shuo-Ting Yan, Richard Behringer, Randy Johnson, David Chiang, Na Li, Ross Poché, Patrick Ellinor, and Wouter de Laat.

John and Melissa Leach, Matt Hill, Tanner Monroe, Allie Shrem, José Islas, Gabi Guevara, and Mahdis Rahmani, my great friends, we spend way too much time together talking about life, science and everything in-between. I will remember this time fondly and am grateful for all of your friendships. Special thanks to John and Matt for reading the draft of this dissertation and giving copious detailed suggestions.

Lastly, I thank my parents for their faith and support. Back in 1990, a sunny day of September, they took a picture of me first day I went to elementary school, ever since then, they have been keeping track of every single step I made. And now 25 years later, I still need to call them everyday from thousands miles away to tell them how I did in school. That is some kind of patience only parents can give.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION AND LITERATURE REVIEW

Use of high-throughput sequencing technology has revolutionized our ability to study heart development. On an unprecedented scale, it allows investigators to precisely measure gene expression, as well as to pinpoint *cis*-regulatory elements. Accordingly, investigators are adopting a new perspective—studying developmental processes and disease conditions in a more systematic and unbiased fashion. Therefore, the outcome of such studies is largely dependent on our ability to use bioinformatic tools extracting crucial information from large-scale datasets. In general, we solve a biological problem by following a series of sequential steps: starting with the identification of key elements from individual datasets; then bridging key elements from one dataset to another; next, manipulating one system and observing the consequences; lastly, applying the implication from one system to another and testing the hypothesis experimentally.

In this dissertation, I will use two examples to explain this thought process (Figure 1.1). First, I will describe the comparison of genomic information between humans and mice. We searched for genetic features causing human cardiac arrhythmias, and tested this hypothesis in a mouse model. Second, I will demonstrate an integrative bioinformatic analysis that reveals gene expression patterns across developmental stages, which leads us to identify key components of heart regeneration.

1.1 Gene expression profiling in cardiac development processes and disease conditions

The Encyclopedia of DNA Elements (ENCODE) project estimated that over 60% of the human genome is transcribed into RNA [33]. Polyadenylated RNA-Sequencing is most commonly used to profile messenger RNA (mRNA) transcription, including

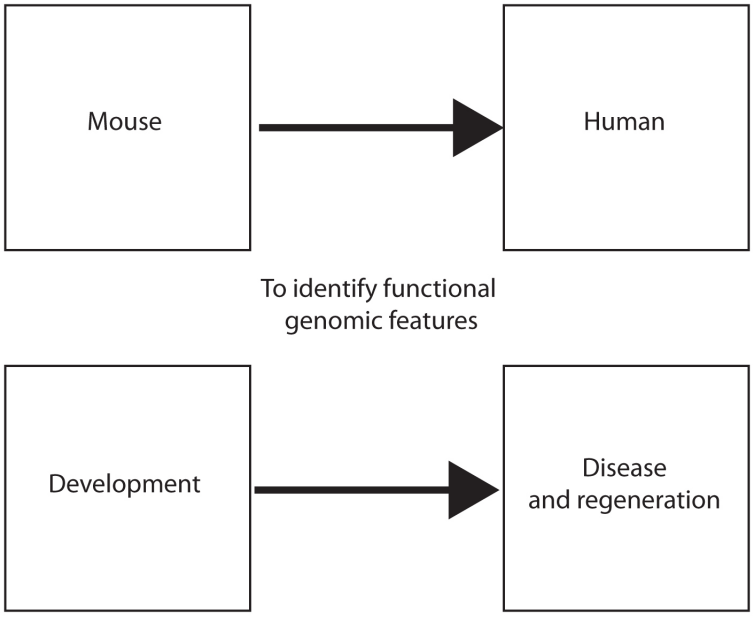


Figure 1.1: Approaches to identify function genomic features. Bridging existing knowledge to unknowns is the key to identify genomic feature.

the majority of the protein coding genes, which cover only 2.9% of the genome [33].

Currently (May 2015), at least 50 RNA-Seq-based studies done by using cardiac tissue, primary cells, or cell lines, are publicly available on Gene Expression Omnibus (GEO) database. These datasets represent different aspects of cardiac development and diseases, such as gene expression profiling of different developmental stages, genetic knockout models, disease models, *in vitro* cardiomyocyte (CM) differentiation, and different subtypes of the heart tissue. These datasets also cover the cardiac transcriptome from a wide range of species, including human (*Homo sapiens*), mouse (*Mus musculus*), zebrafish (*Danio rerio*), rats (*Rattus norvegicus*), rhesus macaque (*Macaca mulatta*), chicken (*Gallus gallus*), sheep (*Ovis aries*) and so on, which will allow us to discover conserved gene expression programs in heart development. The integrative bioinformatic analysis of these individual studies is challenging. For example, it is crit-

ical to select proper methods for normalizing data from different sources. Nonetheless, a comprehensive bioinformatic analysis that integrates individual dataset will provide us valuable information on the dynamic gene transcription process of heart development and disease conditions.

1.2 Temporal gene expression in heart development

Heart formation is well-controlled temporally. In mice, it only takes six days to develop from the primitive streak (mouse embryonic day 6.0, E6.0) to a four-chamber heart (around E12.5) [112]. It requires precise genetic regulation for progenitor cell to properly proliferate and migrate into future heart fields; to form left-right asymmetrical pattern; to determine cell fate and to differentiate into multiple lineages. Emerging evidence indicates that lineage-specific transcription factors (TFs) and signaling molecules orchestrate this drastic morphological change. RNA-Seq datasets taken from snapshots of multiple developmental stages (Table 1.1) summarize the temporal expression pattern of cardiac TFs and signaling molecules. These high-throughput datasets recapitulate the discoveries from decades of elegant work using morphological and gene disruption strategies. More importantly, RNA-Seq datasets will unveil novel candidate regulators that have important functions, yet have been missed by traditional methods.

The best evidence to support the idea that TFs are drivers of robust gene expression and cardiogenesis is using signaling molecules and TFs to progressively differentiate embryonic stem cells (ESCs) to cardiomyocytes (CMs) *in vitro* [91, 59]. The initial efforts to derive cardiomyocytes from murine and human ESCs [112] attempted to reproduce interactions between germ layers through the formation of three-dimensional clusters, termed embryoid bodies, using serum stimulation and suspension culture. Spontaneously contracting CMs typically arise within 4 to 8 days after plating those

Table 1.1: Summary of RNA-Seq datasets from cardiac tissue and primary cells.

Tissue / Cell line	Stage / Cell type	Data references
mouse cardiac progenitors	E4.75	GSE58363 [26]
mouse whole heart	E9.5	GSE63813
	E10.5	GSE56902 [35]
	E11.5	GSE58157 [145]
	E12.5	GSE61209
mouse ventricle	E14.5	[147]
mouse ventricle, CM and CF	E17, P1, P10, P28, P90	GSE49906 [38]
mouse whole heart	P56	GSE49417 [151]
mouse atria	E18.5	GSE52080 [92]
human atria	adult	GSE31999 [52]
human whole heart	adult	GSE49417 [151]
human ventricle	adult ISCH, DCM	GSE57344 [80]
human ventricle	adult ISCH, DCM	GSE55296 [129]
human ventricle	adult ISCH	GSE48166
mouse ventricle and CM	P0, P4, P7, apex resection	GSE64403 [97]
in vitro differentiated CM	day 0, 4, 5, 10	GSE47950 [136]
mouse TRAP *cTnT*-, *Tie2-Cre*	adult TAC	GSE45152 [155]

E: Embryonic day
P: postnatal day
CM: cardiomyocyte
CF: cardiac fibroblast
ISCH: ischemic heart disease
DCM: dilated cardiomyopathy
TRAP: translating ribosome affinity purification
TAC: transaortic constriction

embryoid bodies. Although this strategy is simple and reproducible, the yield of CMs is low. The total number of CMs generated from each mouse embryoid body is about 1-5% of total cells; only 0.1-1% from human ESCs. Cued by genetic principles identified from heart development, Kattman *et al* [59] used Activin and BMP4 to induce dissociated mouse embryoid body cells to become Flk1 and PdgfR-α double positive cells. Furthermore, they treated these cardiogenic mesodermal cells with VEGF and FGF and observed 70% cardiac Troponin T (*cTnT*)-positive cells [136]. The differentiating cells were highly enriched with mesodermal markers (*Mesp1* and *Brachyury*) at day 4, cardiac TFs (*Nkx2-5*, *Tbx5* and *Isl1*) at day 5 and CM-specific genes (*Myh6* and *Myh7*) by day 10. RNA-Seq analysis revealed a close correlation between the expression profiles of these differentiating cells and that observed *in vivo* [136]. Although the differentiated CMs have immature characteristics, for example, *Myh7* is highly expressed at day 10 stage, the protocol still proves to be a promising model for studying CM differentiation and gene regulatory networks.

CM differentiation process does not stop after heart formation. Once the primitive embryonic heart is formed, CMs undergo morphological and physiological maturation. By comparing E17 to adult ventricles, Giudice *et al* [38] showed that extensive gene expression change happen within the first 4 weeks after birth. The expression of genes involved in cell cycle and DNA replication are significantly reduced postnatally, which is consistent with a low proliferation rate and poor regenerative capacity of adult CM [102, 97]. Adult heart transcriptome shows enrichment of genes related to mitochondrial function, fatty-acid metabolism, and oxidation-reduction process, mirrors the metabolic switch from glycolysis to lipid oxidation that occurs after birth. Furthermore, Puente *et al* [103] connected the metabolism pathway to gene program involved in CM cell cycle; demonstrated the oxygen-rich postnatal environment is the upstream signal of CM cell-cycle arrest; and suggested reduction of mitochondrial-

dependent oxidative stress is an important component to ensure CM re-enter cell-cycle in cardiac regeneration.

1.3 Tissue and cell type specific transcriptome

Adult heart is a mixture of multiple cell lineages and differentiated cell types. The non-CMs, which consist of fibroblast, endothelial cells, and smooth muscle cells, are derived from multiple sources: epicardium, endocardium, and neural crest. The flow-cytometry-based studies done in rat and mouse suggest the heart is composed of 50-70% non-CM and 30-50% CM. Since RNA extracted from whole heart represents the average transcript level from both CMs and non-CMs, measure of transcriptome from whole tissue will result in loss of tissue-specific information or mistakenly assign gene expression from one cell type to another. Therefore, it is important to extract tissue- or cell type-specific transcriptome in order to address how different cell types coordinate with each other while maintaining their distinctive identities. Three strategies are commonly used:

(a) Cellular dissociation. The majority of non-CMs are cardiac fibroblasts (CFs) that are often isolated by differential attachment assay, a method that takes advantage of the shorter attachment time of CFs after *Langendorff* perfusion and cellular dissociation. Giudice *et al* [38] used differential attachment method to separate CMs and CFs and performed RNA-Seq across multiple postnatal stages. Interestingly, transcriptome analysis demonstrated a dramatic increase in mature CM gene expression signatures during postnatal stages. However, CF enriched transcripts maintained constant ratios during development, suggesting a different maturation process between CM and CF [38]. To separate embryonic epicardial cells, Xiao *et al* [147] placed E13.5 ventricles in a type I collagen-coated dish for 24 hours. The epicardial cells migrated and formed a monolayer colony on the dish. RNA-Seq analysis revealed that the murine epicar-

dial cells are expressing mesenchymal stem cell surface markers (*Sca1*, *Cd44*, *Cd90*, *Cd49e*, *Cd51*, and *Cd81*) [20]. Furthermore, these cells have differentiation potential for endothelial, smooth muscle and fibroblast lineages.

(b) Regional dissection. Cardiac conduction system is essentially myocardial [112]. The sinoatrial node (SAN) is the primary pacemaking component that generates the electric impulse. In the mature heart, the SAN is located at the junction of the superior caval vein and right atrium. The electric impulse propagates rapidly through right atrial muscle and slowly through the atrioventricular (AV) node, located in the bottom of the right atrium [112]. Given the anatomical differences, left and right atria have distinct susceptibilities towards developing arrhythmias, with left atrial arrhythmias are more common. Gene expression differences established during development are maintained in adulthood, examples include *PITX2* enriched in the left atria and *BMP10* in the right atria [52]. Vedantham *et al* [133] used laser capture micro-dissection (LCM) to isolate RNA from pacemaker cells for RNA-Seq. Differential gene expression and network analysis identified that the TF *Isl1* is active in developing pacemaker cells with a major role in SAN establishment.

(c) Genetic tagging. While cellular dissociation and regional dissection methods are extremely useful methods to isolate a specific cell population, both methods have their shortcomings: the former alters physical environment that maintained *in vivo*; the latter hardly allows a clean separation of different cell types. In contrast, genetic tagging methods enriches RNA by lineage-specific labeling and pull-down, without disrupting native homeostasis of the cells. Zhou *et al* [155] generated a mouse line with lineage-selective Cre actives the expression of GFP fused ribosomal L10a protein, allowing isolation of genetically tagged ribosomes (translating ribosome affinity purification; TRAP). By using *Tie2-Cre* and *cTnt-Cre*, endothelial and CM-specific transcripts were enriched in the substrates of GFP-immunoprecipitation, respectively.

Besides the listing common strategies, newly developed technologies, such as single-cell RNA-Seq [46, 55] and *in situ* RNA-Seq [67] enable massive parallel transcriptome analysis in single-cell resolution, will be extremely useful to advance our understanding in cellular and spatial organization of gene expression programs in heart development and diseases.

1.4 Transcriptome analysis in disease conditions

Heart failure, caused by massive CM loss, is the number one cause of death worldwide. Two of the most common subtypes of heart failure are ischemic heart disease (ISCH) and dilated cardiomyopathy (DCM). Although these two subtypes lead to similar outcomes, they have completely divergent functional and phenotypic progression. Hence, the distinct molecular characterization of these disease conditions is the key for understanding the progression of heart failure and offering therapeutic solutions.

Multiple groups have reported gene expression profiling studies in human ISCH and DCM patients, and showed distinct gene expression signatures between ISCH and DCM [65, 129, 80]. However, interpretation of human data is challenging, due to limited sample size, proper gender and age controls, ethnic backgrounds, and varied co-morbidities. Thus multiple heart injury models have been developed for studying pathology of the disease and regeneration. Through ligation of the left anterior descending coronary artery, myocardium infarction (MI) achieves acute infarction in about two weeks. The expression of cardiac stress genes (*Nppa*, *Nppb*, *Myh6*) and fibrosis genes (*Col1a1*, *Postn* and *Tgfβ1*) are increased in the failing hearts two weeks after MI [98].

Transaortic constriction (TAC) as a model of heart failure, mimics the pressure overload-induced cardiac hypertrophy at one week, and dilated cardiomyopathy at 8 week. Lee *et al* [68] collected left ventricle RNA from hypertrophy and DCM samples,

and found increased expression of genes involved in vascular remodeling, inflammatory response, and fibrosis. They also discovered differentially expressed isoforms in TAC conditions compared to sham controls.

Although the transcriptome analysis of murine failing hearts provides valuable insights in molecular changes of heart remodeling, two major shortcomings are noticeable. Firstly, the failing hearts are mostly fibrotic tissue, limiting our ability to dissect specific gene expression profile from individual cell types. Thus cell-type-specific enrichment, such as lineage specific RNA pull-down methods should be considered. Secondly, more intermediate stages are needed to capture molecular changes before significant cellular remodeling happens.

The mammalian heart maintains regeneration capability in a short window after birth. Neonatal mice can fully regenerate their hearts following ventricular apex resection [102]. Genetic mapping showed the regenerating tissue are from existing CM, instead of a resident stem cell or progenitor population. O'Meara *et al* [97] compared neonatal ventricle, as well as purified CM following apex resection, and found that CM de-differentiated to a primitive stage. The increased gene expression program included reactivation of cell cycle and developmental genes.

Beside heart failure, cardiac arrhythmia is another common disease in clinic. The molecular mechanisms underlying the early development of atrial fibrillation (AF) remains unclear. Chiang *et al* [18] performed RNA-Seq using right atrial appendage samples from paroxysmal AF patients and found the expression of inflammatory related signaling pathway were increased. Furthermore, they discussed the potential involvement of the gonadotropin releasing hormone receptor and p53 pathways in AF pathogenesis. However, the study has limited sample size and only right atrial samples were obtained.

1.5 Dark matter of the transcriptome

Analysis of polyadenylated enriched RNA-Seq is largely dependent on bioinformatic tools and genome annotations. Existing RNA-Seq analysis tools can be grouped into two classes: transcript-based assembly and differential expression analysis tools and exon-based quantification tools. First group included assembly tools like TopHat and Cufflinks [62], which assemble RNA-Seq reads to individual transcripts based on splice junctions from either reference annotation or using a *de novo* assembly algorithm. The advantage of this method is that the abundance of different isoforms is reported. In particular, the *de novo* assembly algorithm enables the detection of novel transcripts and alternative splicing events. However, this method requires a large number of sequencing reads to accurately predict splice junctions. By contrast, the tools in the second group, such as HTSeq [4], count the raw reads based on individual exons. The quantification of gene expression is reported based on sum of all the exons in a particular gene, not transcript. Ambiguity arises when multiple gene bodies are overlaid with each other. The most common solutions are either to arbitrarily assign the reads to either gene, or to simply discard the reads. Therefore approximately 50% of the reads from a given RNA-Seq experiment will not be annotated by these tools.

What makes up the remaining transcriptome? Long non-coding RNAs (lncRNAs) are RNA transcripts longer than 200 nt, often poly-adenylated and 5'-capped, with lower expression levels and less conservation compared to protein-coding RNAs [42]. Interestingly, large-scale RNAi screening showed that lncRNA function in ESC differentiation [43]. Klattenhoff *et al* [64] and Grote *et al* [40] identified lncRNAs necessary for specifying the cardiac lineage. Specifically depletion of *Braveheart*, a lncRNA which regulates *Mesp1* through physically interacting with component of polycomb-repressive complex (PRC), resulted in CM differentiation defects *in vitro*. Similarly,

Fendrr, a lncRNA which represses PRC, and promotes cardiac-specific gene expression. In addition, Han *et al* [45] demonstrated *Myheart*, an anti-sense transcript of *Myh7*, predominately expressed in the nucleus, prevents chromatin from remodeling following cardiac hypertrophy. Thus these studies indicate the importance of lncRNA during cardiac lineage commitment and disease conditions. However, all of these lncRNAs are not annotated in the current version of the mouse genome. Only through deep RNA-Seq can these lncRNAs be discovered.

The enhancer RNAs (eRNAs), present a group of novel and relative unstudied short non-coding RNAs. Wu *et al* [145] identified over 3,000 eRNA expression in the mouse heart, which globally correlate with tissue-specific enhancer activity. Given that a large amount of eRNA do not contain poly-adenylated tails, poly-adenylated enriched RNA-Seq can therefore only detect a small proportion of the total eRNAs. One solution to discover novel eRNAs is to use total RNA-Seq. However, the sensitivity of total RNA-Seq is largely affected by the abundant ribosomal RNA population. A better strategy is to use global nuclear run-on sequencing (GRO-Seq), which captures nascent RNA transcripts. Luo *et al* demonstrated that *NF-κB* associated eRNAs are turned on in response to TNF-α stimulation in an immortalized adult ventricular CM cell line, AC16. This suggests a potential role of eRNAs in the *cis*-regulatory network. Therefore, GRO-Seq is a robust technique which may be applied towards investigating the transcriptome dynamics occurring between transient stages, such as a intermediate stages throughout heart regeneration.

Small non-coding RNAs (< 200 nt), including microRNAs and small nucleolar RNAs (snoRNAs), are captured using size-selected RNA-Seq. Mature microRNA sequences are highly conserved throughout all metazoans. MicroRNAs expressed in the heart has been widely studied. Known cardiac enriched microRNAs include miR-1, 133, 206, 208, 499, 17-92 and 106b [139]. By contrast, the function of snoRNAs

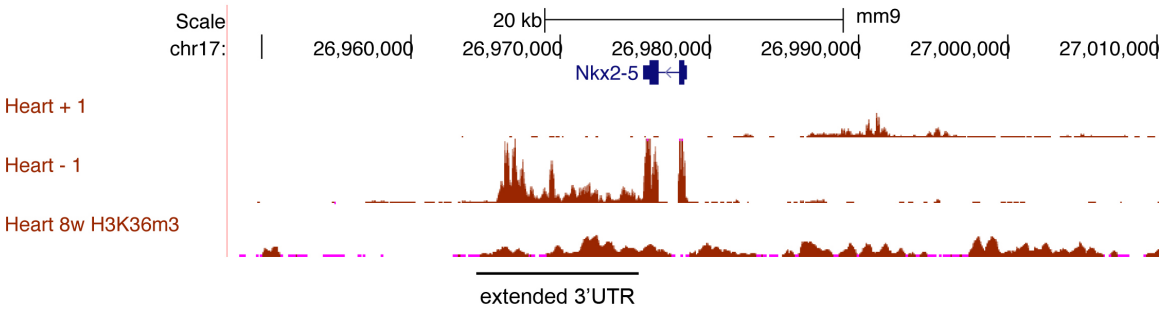


Figure 1.2: Extended 3' untranslated region in *Nkx2-5*. RNA-Seq were performed using 8-week-old mouse heart. The forward and reverse strands are labeled as '+' and '−', respectively. H3K36me3 ChIP-Seq from same tissue indicates active transcription.

in heart is largely unknown. Our preliminary small RNA-Seq dataset from the heart indicates that only 40% of reads are annotated. The functional analysis of unannotated cardiac small RNA represents an unexplored frontier in this field.

Alternative poly-adenylation (APA) is a pervasive mechanism which plays a part in the regulation of most mammalian genes. The role of APA-mediated gene regulation in heart development and diseases is just beginning to be appreciated. Xia *et al* [146] used a novel bioinformatics-based algorithm to identify *de novo* APA from the Cancer Genome Atlas (TCGA) RNA-Seq datasets, and found that most APA-regulated genes have short 3'-UTRs. Since 3'-UTRs have been known harboring microRNA seed sites, shortened 3'-UTRs indicate the potential loss of microRNA-mediated repression. Conversely, while examining through RNA-Seq data generated from heart tissue, we found that a master cardiac regulator gene, *Nkx2-5*, has an unannotated transcript with an extended 3'-UTR (Figure 1.2). Thus, we believe that a genome-wide categorization of APA across developmental stages and disease states will reveal novel mechanisms of cardiac gene regulation.

Table 1.2: Histone modification marks. Core set of five chromatin modification marks.

Chromatin modification	Associated chromatin state
histone H3 lysine 4 trimethylation (H3K4me3)	promoter regions
H3 lysine 4 monomethylation (H3K4me1)	enhancer activity
H3 lysine 27 acetylation (H3K27ac)	enhancer activity
H3 lysine 36 trimethylation (H3K36me3)	transcribed regions
H3 lysine 27 trimethylation (H3K27me3)	Polycomb repression
H3 lysine 9 trimethylation (H3K9me3)	heterochromatin regions

1.6 Non-coding regulatory sequences

Before human genome project initiated, the estimated number of genes in human genome ranged from 50,000 genes to as many as 140,000. With the completion of first draft of human genome, startling finding was the number of human genes is significantly fewer than previously estimated. In fact, both human genome and roundworm (*Caenorhabditis elegans*) genome contain around 20,000 genes. Thus the complexity of mammalian genomes does not related to their number of genes, instead 95% of the mammlian genomes are non-coding regions [23, 110]. These non-coding regions often harbor regulatory elements, such as enhancers that are sequences bound by TFs and promoter gene expression via three-dimensional chromatin structures. With the development of high-throughput sequencing technology, millions of potential regulatory elements are predicted by TFs binding sites and histone modification signatures (Table 1.2). Furthermore, chromatin conformation capture technology facilitates our understanding of the complex and versatile mechanism underlying gene-enhancer selection and the landscape of regulatory elements across the genome.

1.7 Transcription factors

Transcription factors (TFs) are evolutionarily conserved and play key roles in cell-fate specification [51]. Pioneered by Yamanaka [127, 126], using only several ESC-specific TFs (*Oct3/4*, *Sox2*, *Klf4*, and *c-Myc*) can induce differentiated cells to an embryonic stem cell-like state. Similarly, by the addition of cardiac TFs, adult somatic cells can be directly reprogrammed into cardiomyocyte-like cells, both *in vitro* and *in vivo* [106, 105, 54]. The same concept is widely applied to non-CM cell types, for example, to reprogram fibroblast to endothelial-like cells [85].

The key role of TFs is to bind DNA elements and drive spatiotemporal pattern of gene expression. ChIP-Seq, or chromatin-immunoprecipitation followed by deep sequencing, has been widely adopted for studying genome-wide occupancy of TFs, as well as histone modifications. So far, multiple ChIP-Seq datasets of key cardiac TFs from murine hearts and cardiac related cell lines are available through GEO database (see review: [137]). Interestingly, many of binding sites from different TFs are co-localized together, especially in the open chromatin regions that are marked by Dnas I hypersensitive sites (DHS) and enhancer regions marked by H3K4me1, H3K27ac and p300 ChIP-Seq. Furthermore, one TF can control expression of other TFs via *cis*-regulation, indicating a well-organized transcriptional regulatory circuitry underlies heart development [13].

Using bioinformatics algorithms, enriched TF binding peaks are detected by comparing either to an input dataset or locally over sequencing background. Size of the ChIP-Seq peaks is commonly around 200-500bp. To assess the quality of ChIP-Seq datasets, motif analysis is often required. Consensus sequence of TF binding site, or motif, is expected to be centered and significantly enriched in ChIP-Seq peaks. Such consensus motifs were acquired both experimentally and computationally. Weirauch

et al [142] broadly sampled DNA binding motifs from multiple eukaryotic clades and determined DNA binding preferences of over 1000 TFs using an oligo array. Closely related TFs almost always have similar DNA binding motifs. Moreover, multiple motifs are often enriched in one ChIP-Seq experiment, indicating potential co-factors are binding to same locations with the factor of interest.

1.8 DnaseI hypersensitive sites and enhancers

DHS, denoting accessible chromatin regions, are the indicators of open chromatin. Regulators associated to these regions include TFs, histone modification marks, chromatin insulators, cohesin complex, and lamina-associated domains. High-coverage DHS-Seq, or digital genomic foot-printing (DGF) [110], uses computational methods to search for tissue-specific TF binding sites. The epigenetic roadmap project provides DHS-Seq from human and mouse heart, which are valuable references for genome-wide epigenetic states in heart. However, DHS-Seq is technically challenging —its protocol requires large amount of tissue and involves low-efficient steps. Recently, a transposase-based assay, transposase-accessible chromatin using sequencing (ATAC-Seq), was developed for capturing open chromatin sites with easier protocol and higher sensitivity.

Enhancers are distal regulatory elements that can activate tissue-specific gene expression and are abundant throughout mammalian genome [96]. Using H3K27ac as a mark of active enhancers, Nord *et al* [96] identified over 90,000 enhancers across three different mouse tissues. Most of these enhancers exhibited tightly restricted temporal and spatial activity. In addition, the transgenic mouse assay exhibited 12/18 (67%) of tested enhancer drove reproducible expression pattern *in vivo*. Using an *in vitro* differential protocol, Wamstad *et al* [136] identified highly correlated patterns between gene expression and epigenetic modification. They further identified stage-specific

distal enhancer elements and found that binding motifs of driver TFs were enriched in these enhancers.

Overall, high-throughput sequence technology opens up a new era in biology research. Publication of a dataset no longer means the end of the study, but an addition of valuable resource to the community. Empowered by integrated bioinformatic analysis, high-quality datasets will be mined again and again. The discovery from highly integrated analysis will advance our understanding of origins of the heart, will unveil the cause of diseases, and will shed light on novel therapeutic targets.

2. *PITX2* DIRECTLY TARGETS ION TRANSPORT AND INTERCALATED DISC GENES*

Atrial fibrillation (AF), the most prevalent adult arrhythmia in humans, results in an increased risk of stroke, dementia, and heart failure [10]. Electric impulses that are critical for a coordinated, physiological heartbeat originate in the sinoatrial node and are transduced from the pacemaker region into both atria. In AF, fibrillatory atrial impulses override normal conduction pathways with resulting irregular ventricular conduction.

Because of its clinical significance, great efforts have been expended to investigate the genetic underpinnings of AF using unbiased genome-wide approaches such as genome-wide association study (GWAS). Multiple studies have revealed a single-nucleotide variant on human 4q25 that was associated with familial AF [84]. Patients with the 4q25 variant exhibited early onset AF that was independent from known AF risk factors. The 4q25 single-nucleotide variant also had prognostic value because patients with this variant are prone to AF recurrence after ablation therapy [53]. The gene-poor 4q25 region harbors the *Pitx2* homeobox gene, which has been implicated in AF predisposition using mouse models [138, 19, 63, 2].

Pitx2, a transcription factor that controls transcription of numerous target genes, is located in proximity to sequence variants in the human genome that are commonly associated with atrial fibrillation (AF). Although previous work focused on *Pitx2* during development, the current study investigated *Pitx2* in postnatal heart. It was unknown previously whether Pitx2 has distinct postnatal and developmental functions.

*Reprinted with permission from "Pitx2, an atrial fibrillation predisposition gene, directly regulates ion transport and intercalated disc genes" by Y. Tao, M. Zhang, L. Li, Y. Bai, Y. Zhou, A. M. Moon, H. J. Kaminski, J. F. Martin. *Circulation: Cardiovascular Genetics*, 1(7):23-32, Feb 2014. Wolters Kluwer Health Lippincott Williams & Wilkins©. No modifications will be permitted.

The *Pitx2* genes were removed from hearts of postnatal mice. Unstressed adult *Pitx2* mutant mice display arrhythmias called sinus node dysfunction, an AF risk factor in human patients. To uncover target genes that are regulated by Pitx2, we used a genome-wide approach to identify all cardiac genes that are regulated by Pitx2. Pitx2 target genes encoded cell junction proteins, ion channels, and critical transcriptional regulators. Importantly, many Pitx2 target genes have been implicated previously in human AF. Other studies in adult *Pitx2* mutant mice revealed structural remodeling in the heart characteristic of human patients with AF. Our findings provide new mechanistic insight into AF, revealing that *Pitx2* has genetically separable postnatal and developmental functions and unveiling direct Pitx2 target genes that include channel and calcium handling genes, as well as genes that stabilize the cellular structure in postnatal atrium. Because Pitx2 regulates many of these genes in the postnatal heart, it is conceivable that drugs can be developed to modulate the molecular interaction between Pitx2 and its target genes. Our findings that many important *Pitx2*-regulated events occur postnatally also strengthens the likelihood that *Pitx2*-mediated AF may be treatable in the future.

2.1 AF and associated arrhythmias

AF may result from new, pathological sources of electric impulses. For example, many cases of ectopic electric activity originate in the pulmonary vein [44]. Other sites of ectopy include the left atrial posterior wall, superior vena cava, interatrial septum, crista terminalis, and coronary sinus myocardium [58, 74]. In addition to ectopy, other causes of AF involve atrial myopathy that disrupts normal atrial conduction and promotes re-entrant circuits. One common example of AF secondary to myopathy is fibrosis that in some cases may be because of elevated transforming growth factor-β signaling [34].

Work from the Framingham study has shown that patients with PR interval prolongation, also called first-degree atrioventricular block, often develop AF [17]. In addition to atrioventricular block, sinus node dysfunction (SND) is also an AF risk factor in human patients [111]. Notably, progression to higher grade arrhythmias with time is also common, reflecting the importance of aging in arrhythmogenesis. Although the mechanistic connection among SND, PR interval prolongation, and AF are poorly understood, all three conditions may involve an atrial myopathy with defective atrial impulse conduction [69].

Predisposition to AF may result from a developmental defect that results in an adult heart with subclinical abnormalities that subsequently manifest as overt disease after environmental insults or aging. Alternatively, postnatal homeostatic genes may be required to maintain normal tissue structure and physiology. Small changes in homeostatic gene level may result in subclinical disease until an AF-inducing stress is encountered.

2.2 *Pitx2* and predisposition to AF

Pitx2 encodes 3 isoforms, *Pitx2a*, *Pitx2b*, and *Pitx2c*, that are generated by alternative splicing and dual promoter usage [138]. *Pitx2c* is generated via an intergenic promoter, whereas the *Pitx2a* and *Pitx2b* isoforms that are generated by alternative splicing use an upstream 5' flanking promoter. The Pitx2c isoform is expressed on the left side of the embryo, whereas the *Pitx2a* and *Pitx2b* isoforms are expressed symmetrically in the head within the eyes and craniofacial structures [138]. In mouse, *Pitx2c* expression continues in the postnatal atrium, whereas human *PITX2C* is also the predominant isoform in the left atrium [138, 63, 52]. Studies in isoform-specific knockout mice in our laboratory also revealed that the *Pitx2c* isoform is the dominant isoform in determining left right asymmetry during development [77]. *Pitx2* haploinsufficient

(*Pitx2null*$^{+/-}$) adult mice and *Pitx2c*$^{+/-}$ adult mice were prone to AF when challenged by programmed stimulation, indicating that reduced Pitx2c levels during development create an arrhythmogenic substrate [138, 63]. Notably, Pitx2 levels are also decreased in the atria of human patients with AF [19].

Previous experiments indicated that sinoatrial node genes were expanded in left superior caval vein and left atrium of *Pitx2*$^{null/+}$ and *Pitx2*$^{null/null}$ mutant embryos, indicating a developmental defect [138]. Optical mapping experiments showed *Pitx2* conditional mutant embryos had a functional left-sided sinoatrial node that could override normal atrial cardiomyocyte depolarization [2]. Furthermore, *Pitx2c*heterozygous adult mice had shortened action potential duration without fibrosis or structural defects, suggesting an electrophysiological mechanism for arrhythmogenesis in *Pitx2c* germline mutants [63].

It is unknown whether *Pitx2* has a postnatal homeostatic function. To study *Pitx2* postnatal function, we generated a *Pitx2* conditional knockout (CKO) mouse line that deletes Pitx2 in postnatal atrium. The adult Pitx2 CKO mice had abnormal electrocardiography with irregular R–R interval and low-voltage P waves, indicating SND with impaired atrial conduction. A genome-wide search for *Pitx2* targets by ChIP-Seq and microarray assays revealed genes encoding cell adhesion or cell junction proteins, ion channels, and transcription factors. Many *Pitx2* target genes are AF risk genes identified in human GWAS [84]. Using immunofluoresence and transmission electron microscopy (TEM), we obtained evidence for structural remodeling of the intercalated disc (ID), the structure that mediates cardiomyocyte electromechanical coupling [8]. Our data indicate a postnatal role for Pitx2 in AF predisposition and uncover novel Pitx2 target genes that provide new mechanistic pathogeneses for AF.

2.3 Results

2.3.1 Generation of Pitx2 CKO mice

To generate a *Pitx2* conditional loss of function (*Pitx2* CKO) allele that deletes all *Pitx2* isoforms in postnatal atrium, mice with the *MCK-Cre* driver were crossed to mice bearing the *Pitx2* conditional null (Flox) allele (Figure 2.1A). *MCK-Cre* directs Cre activity in skeletal and cardiac muscle with a perinatal onset around birth [156, 14]. We examined Cre activity using the *R26R LacZ* reporter. MCK-Cre activity initiates in atria after birth and within ventricles in a mosaic pattern after E15.5 (Figure 2.1B). Because *Pitx2* is predominantly expressed in left atrium at these fetal stages and in adult, *MCK-Cre* is a valuable tool to address *Pitx2* function in postnatal left atrium [138, 77]. Both immunostaining and real-time reverse transcription PCR revealed that *Pitx2* was efficiently deleted in neonatal left atrium (Figure 2.1C&D). Although *Pitx2* is also inactivated in skeletal muscle, we did not detect skeletal muscle phenotypes in *Pitx2* CKO mice perhaps because of overlapping function with *Pitx3* [71].

2.3.2 Pitx2 CKO mice have abnormal cardiac conduction

We implanted telemetry transmitters into control and *Pitx2* CKO adult mice and collected resting electrocardiographic data in awake mice. Compared with normal sinus rhythm in controls (Figure 2.2A), ECG tracing for *Pitx2* CKO mice shows abnormal heart rate with irregular R–R intervals and low-voltage P waves (Figure 2.2B; Table 2.1). This phenotype was observed in all *Pitx2* CKO mice studied by resting telemetry ECG but none of the controls (Table 2.2). Because sinus node sets heart rhythm and P wave represents atrial depolarization, these phenotypes indicate SND with impaired atrial conduction. Importantly, SND is closely associated with AF in human patients [17].

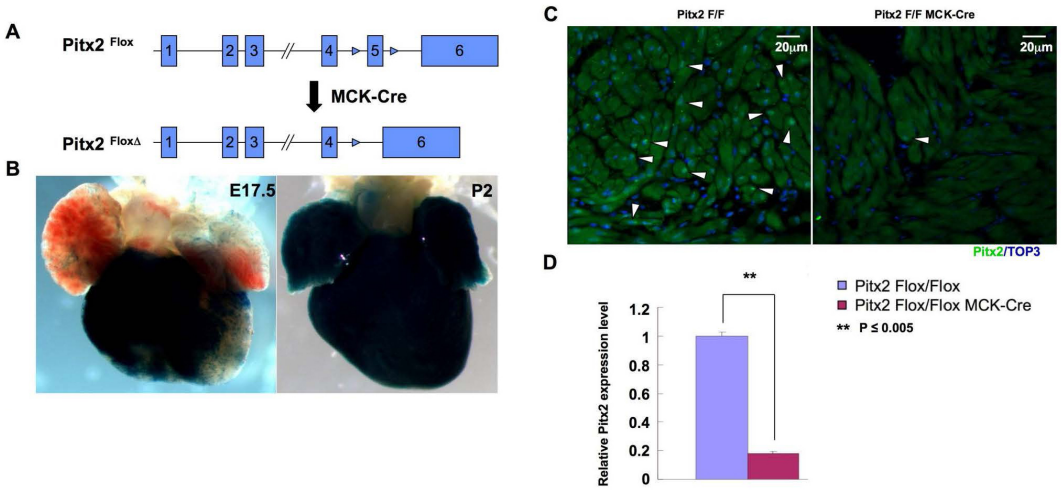


Figure 2.1: Generation of the *Pitx2* conditional knockout mice. (A) Strategy for the generation of the *Pitx2* conditional knockout allele. *Pitx2* Flox allele has two LoxP sites flanking exon 5, which encodes the N-terminus of DNA binding homeodomain. Upon the activation of MCK-Cre, recombination between the two LoxP sites results in the deletion of exon 5, thus generate the Pitx2 conditional knockout allele. (B) Expression of MCK-Cre in the heart. *MCK-Cre* mice were mated to *R26R* mice to generate progenies with the genotype *MCK-Cre R26R* which express β-Galactosidase (β-Gal) in the *MCK-Cre* expression domain. Hearts of embryos and pups from the mating were collected and stained for β-Gal. Images of whole mount E17.5 *MCK-Cre R26R* embryonic heart and P2 heart are shown. In the E17.5 *MCK-Cre R26R* heart, β-Gal staining was detected in the ventricles and was barely detected in the atria (left panel). In P2 heart, β-Gal staining was detected in the whole heart (right panel). (C) Immunofluorescence staining for Pitx2 in the postnatal heart cells show expression of Pitx2 in the nuclei of the control (*Pitx2* $^{Flox/Flox}$) heart cells and is nearly lost in Pitx2 CKO heart. Nuclei of the heart cells were also labeled by TO-PRO-3 (TOP3). Images show the staining in the left atria of the 3-month-old control (*Pitx2* $^{Flox/Flox}$) and *Pitx2* CKO (*Pitx2* $^{Flox/Flox}$ *MCK-Cre*) mice. (D) Real time RT-PCR assay using primers spanning exon 5 of *Pitx2* transcripts shows *Pitx2* expression level decreased over 80% in the left atrium of P1 Pitx2 CKO heart. Values and error bars represent mean and standard deviation ($n = 3$). Figure credit: Ye Tao

Table 2.1: Abnormal R-R intervals in *Pitx2* CKO mice. R-R interval were sampled from uninterrupted one minute ECG tracing of control (*Pitx2* $^{Flox/Flox}$) and *Pitx2* CKO (*Pitx2* $^{Flox/Flox}$ *MCK-Cre*) mice. Mean and standard deviation of R-R interval from each mouse were listed. Two-sided, unpaired Student's t-test was performed for mean and standard deviation values of R-R interval; no significant difference was detected of mean R-R interval value between controls and mutants (p = 0.543); standard deviation of R-R interval from mutant mice were significantly greater than that of control mice (p = 0.003). ms: millisecond. SD: Standard Deviation.
p = 0.543 ** p = 0.003

Animals	Age	# of heart beats	R-R interval (ms)	
			Mean (*)	SD (**)
Ctrl 1	2-4 months	554	105	9
Ctrl 2		392	146	10
Ctrl 3		467	123	10
Ctrl 4		447	127	10
Ctrl 5	7 months	420	142	10
CKO 1	2-4 months	488	117	30
CKO 2		408	140	24
CKO 3		420	137	21
CKO 4		485	118	18
CKO 5		505	114	16
CKO 5		585	98	29
CKO 6	7-8 months	513	113	16
CKO 7		420	137	45

Table 2.2: Mice tested by telemetry ECG. The table summarizes genotype, age, number and phenotype of mice tested by telemetry ECG.

Genotype	Age	# of Animals	Phenotype
Pitx2 $^{Flox/Flox}$	2-4 months	4	Normal sinus rhythm (4/4)
Pitx2 $^{Flox/Flox}$	7-8 months	1	Normal sinus rhythm (1/1)
Pitx2 $^{Flox/Flox}$	2-4 months	5	Sinus node dysfunction with impaired atrial conduction (5/5)
Pitx2 $^{Flox/Flox}$	7-8 months	2	Sinus node dysfunction with impaired atrial conduction (2/2)

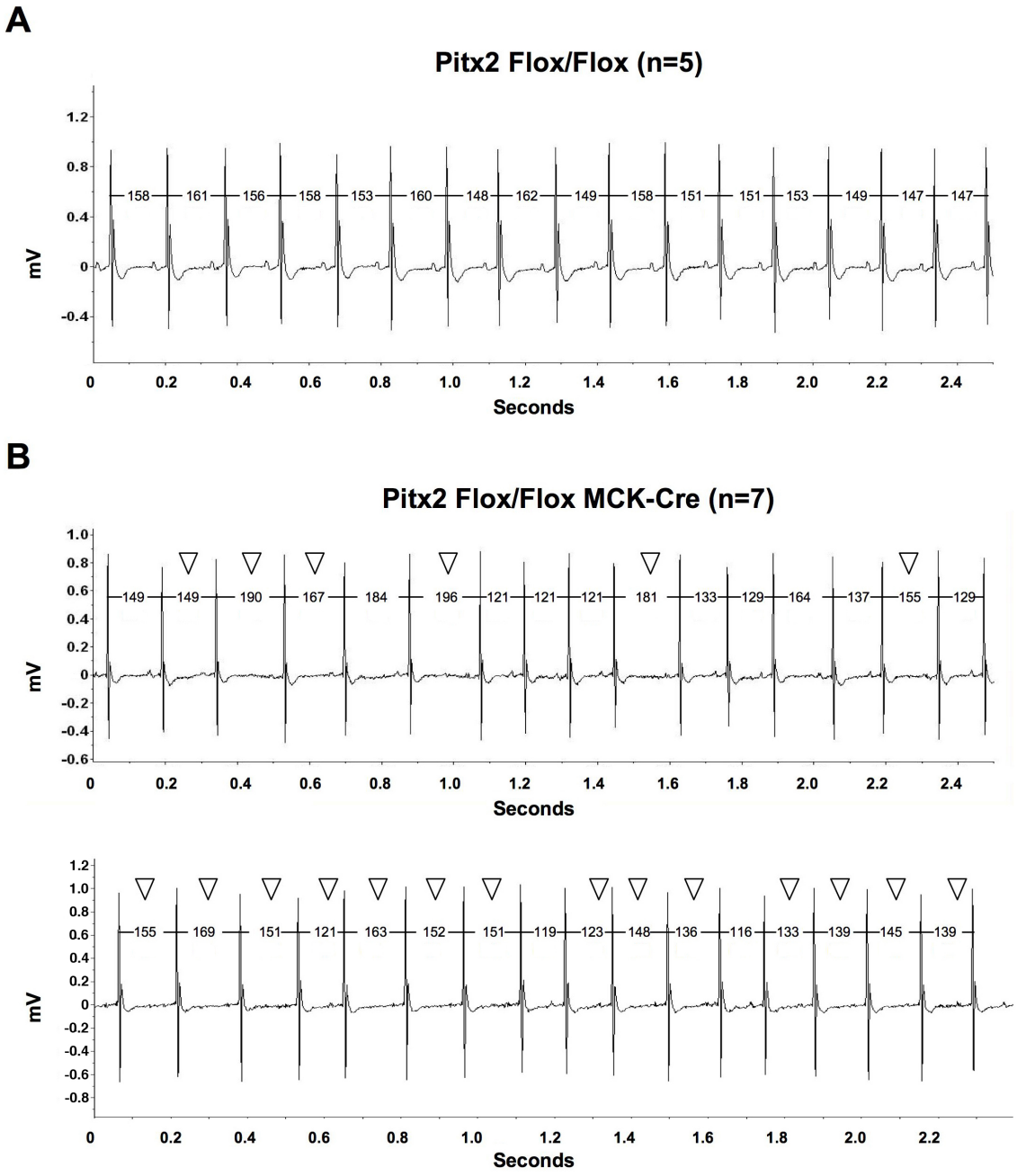
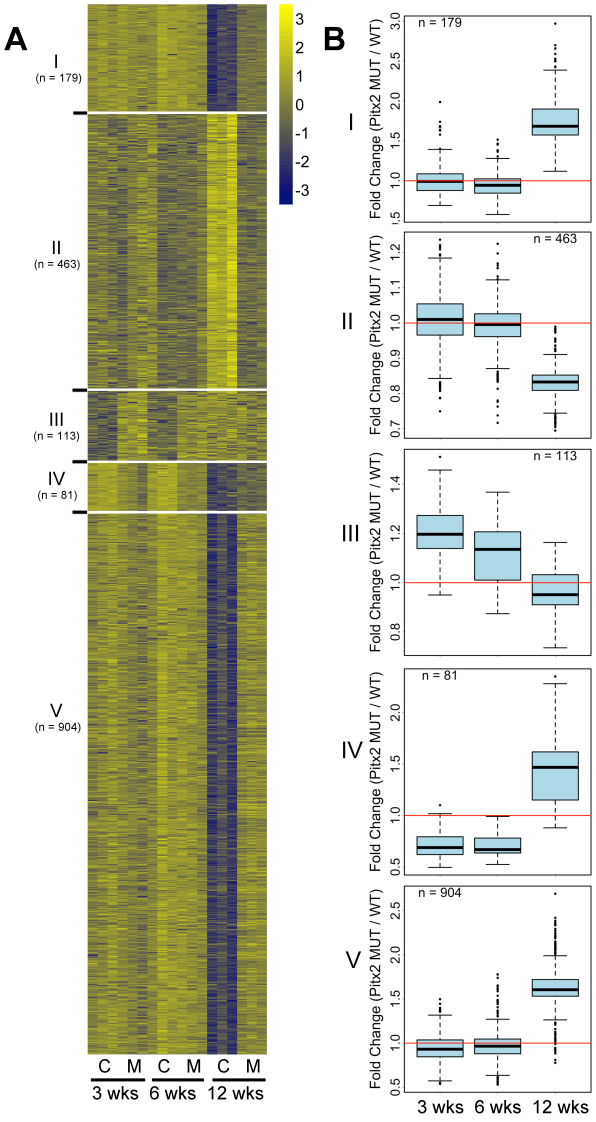


Figure 2.2: *Pitx2* conditional knock out mice have abnormal heart function. Sinus node dysfunction with impaired atrial conduction were observed in *Pitx2* conditional knock out mice. By telemetry ECG tracing in awake mice, heart function of the adult control (*Pitx2* $^{Flox/Flox}$) and *Pitx2* CKO (*Pitx2* $^{Flox/Flox}$ *MCK-Cre*) mice were tested. (A) In control mice, normal sinus rhythm was observed. (B) In the Pitx2 CKO mice, mutant heart function including irregular R-R intervals and low voltage of P waves (open arrowheads) were observed. Tracings from one control and two Pitx2 CKO mice are shown. All mice shown were 3-month-old. The unit of R-R interval is millisecond (ms). Figure credit: Ye Tao

2.3.3 Unbiased discovery of Pitx2 regulated target genes

We performed microarray transcriptional profiling on 3-, 6-, and 12-week-old whole hearts from *Pitx2* CKO and control mice (Figure 2.4). Most genes were upregulated in *Pitx2* CKO mutant hearts, suggesting that Pitx2 is a transcriptional repressor in the adult heart (Figure 2.3A; Figure 2.5). We performed clustering analysis on all differentially expressed genes and identified 5 gene expression categories in *Pitx2* CKO hearts compared with controls. Groups I, IV, and V genes were upregulated significantly in the *Pitx2* CKO at the 12-week stage. In contrast, group III genes were upregulated at the 3- and 6-week stages in *Pitx2* CKO hearts. Group II genes, the only gene class that was downregulated, were reduced at 12 weeks in *Pitx2* CKO hearts (Figure 2.3B). Genes in groups I, IV, and V were involved in cell junction assembly, as well as proliferation and migration (Figure 2.3C).

To investigate Pitx2 direct target genes, a ChIP-Seq assay was performed at the 12-week stage on heart tissue. Overlay of Pitx2 ChIP-Seq data and the 12-week expression profiling revealed genes that are likely direct Pitx2 target genes. Gene ontology analysis of the ChIP-Seq/microarray overlay gene-set indicated that most Pitx2 target genes are involved in cell-cell junction organization and ion channel physiology (Figure 2.6A; Figure 2.7). Another important gene ontology term providing insight into Pitx2-mediated arrhythmias was arrhythmogenic right ventricular cardiomyopathy thought to be a desmosome defect. We categorized Pitx2 direct target genes into 4 main groups based on gene function, including cell junction assembly (group A), ion transport (group B), arrhythmogenic right ventricular cardiomyopathy (group C), and transcriptional regulation (group D, Figure 2.6B). Enriched ChIP-Seq tags (peaks) were identified in potential regulatory chromosomal regions for these targets (Figure 2.9A; Figure 2.8).



C

GO Terms	Z Score
Group I	
adherens junction organization	12.02
alpha-catenin binding	11.35
cell junction assembly	8.50
sarcoplasmic reticulum membrane	8.07
muscle cell proliferation	7.84
regulation of metal ion transport	6.52
contractile fiber part	6.24
Group II	
histone exchange	9.94
beta-amyloid metabolic process	8.16
intracellular organelle	6.75
bone mineralization	6.08
Group III	
mitochondrial nucleoid	6.37
negative regulation of myeloid cell differentiation	4.88
activation of protein kinase activity	4.32
axis specification	4.23
embryonic skeletal system morphogenesis	4.04
Group IV	
core promoter sequence-specific DNA binding	8.08
positive regulation of multicellular organism growth	7.93
neuromuscular junction	6.41
actin filament binding	4.88
Group V	
cytoplasmic part	13.11
structural constituent of ribosome	6.46
regulation of intracellular protein kinase cascade	5.35
regulation of growth	5.30
cell migration involved in heart development	5.23
negative regulation of cellular process	5.16
heart development	5.14

Figure 2.3: Age-related gene expression profiling in *Pitx2* mutants. (A) Clustering analysis of gene expression profile across three time points in control and Pitx2 mutant hearts. Expression patterns of five representative groups are shown in heat map. The color bar represents relative expression level for each gene across different samples. (B) Differential expression between *Pitx2* mutant and control across three time points. Mean value for each group was used to calculate fold change between mutants and controls. Box plots for selected clusters show median (middle line), third and first quartile (top and bottom hinges of the boxes), whiskers and outliers (black dots). (C) Gene ontology analysis of genes in each group. GO terms related to biological process and molecular function categories.

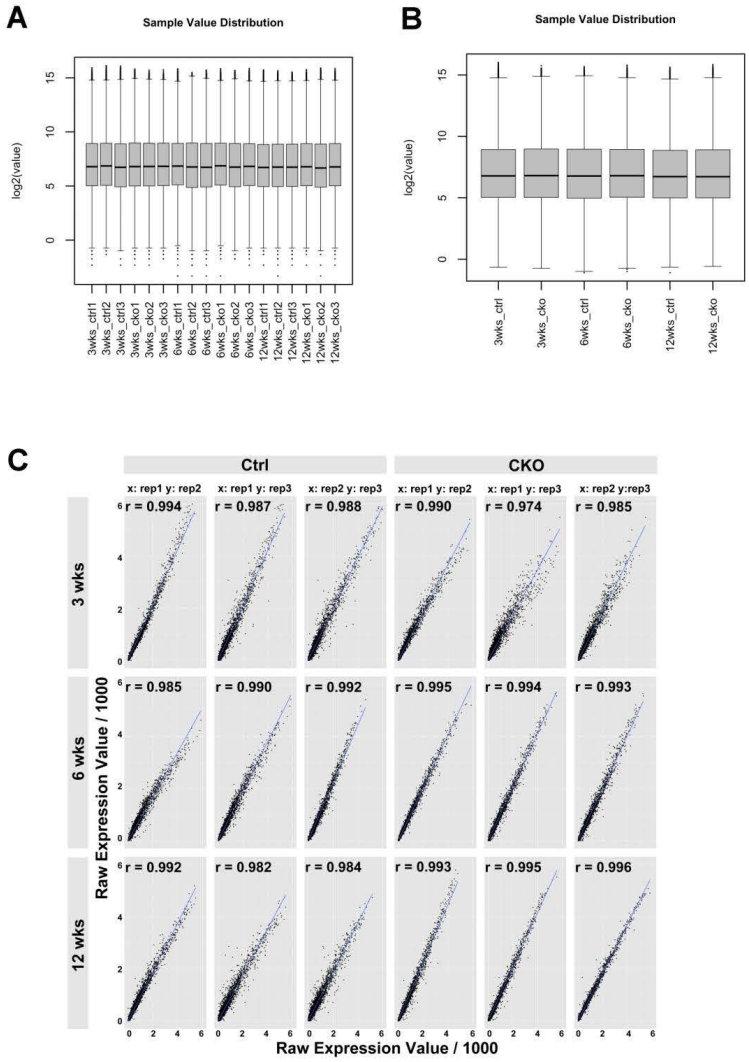


Figure 2.4: Microarray intensity value distribution and reproducibility. (A) Microarray was performed on *Pitx2* control and mutant hearts collected from 3-, 6- and 12-week-old mice. At each time point, three controls and three mutants were collected as biological replicates. Raw signal values were normalized using a logarithmic method. No significant outlier was observed across the samples. (B) Mean of logarithmic raw signal values of biological replicates was used for comparison between groups. No significant outlier was observed across groups. (C) Reproducibility of microarray raw signal across three biological replicates in each group. Each probe was plotted using raw signal value from one biological replicate as *x*-axis value, and that from another biological replicate as *y*-axis value. Total 45,101 points were plotted on each diagram. Linear correlation coefficient was measured, as r given in each diagram. The blue lines on each diagram indicate the line of best fit.

Motif analysis identified the Pitx2-binding element as the most highly enriched in the ChIP-Seq data set validating our experiment (Figure 2.6C) [1]. Other transcription factor binding sites were also enriched in the ChIP-Seq data set, providing further insight into potential Pitx2 co-factors in the adult heart (Figure 2.6C). Among potential Pitx2 cofactors, Nkx2.5 is of great interest because it has been implicated in human AF in GWAS [100].

Our data sets indicated that other genes that have been implicated in human AF through GWAS or other genetic analysis are potential Pitx2 target genes such as *Kcnq1*, *Cav1*, and *Zfhx3* [99, 94, 41, 50, 16, 32, 10]. Other Pitx2 target genes, such as *Cacna1d* and *Tbx20*, have also been implicated directly in AF in mouse models or other human arrhythmias such as prolonged QRS (Table 2.3) [100, 154, 123, 7].

2.3.4 Pitx2 directly regulates genes involved in cell junction assembly, ion transport, and transcriptional regulation

We validated gene expression changes for genes in ChIP-Seq/microarray overlay by real-time reverse transcription PCR. We also evaluated several candidate genes, previously implicated in human arrhythmias, that were unchanged on the microarray. Among these were *Kcnq1*, *Cav1*, *Tbx20*, and *Zfhx3*. In *Pitx2* CKO hearts, we found increases in gene expression levels for ion channel and calcium handling genes such as *Cacna1d*, *Cacna2d2*, *Kcnq1*, *Kcnj11*, *Ryr2*, *Jph2*, and *Atp2a2*. We saw upregulated expression of genes encoding transcriptional regulators Hdac7, Tbx20, and Zfhx3, as well as Cav1, a component of caveolae. Significantly, elevated histone deacetylase (Hdac) activity has been implicated in AF vulnerability, structural remodeling, and fibrosis [78]. Tbx20 is a central component of a gene regulatory network that includes channel genes and other transcriptional regulators [118]. Zfhx3, a homeodomain zinc finger transcription factor that was implicated in AF via GWAS, is also required for

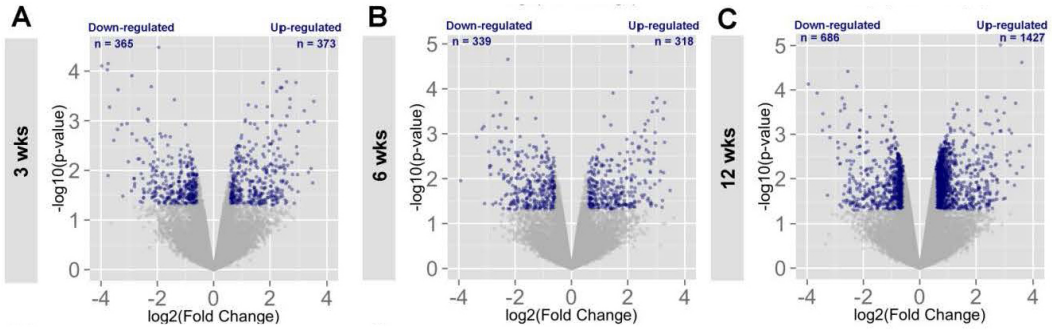


Figure 2.5: Differential expressed genes in each stage. (A-C) *Pitx2* controls and mutants were compared to each other within each stage using linear models. Individual genes were plotted using logarithmic of fold change as x-axis index and -log10 of p-value as y-axis index. Total 45,101 points were plotted on each diagram. Differential expressed genes with a fold change of at least 1.5 and p value < 0.05 were highlighted in blue. The numbers of genes up-regulated or down-regulated in each stage were shown as n.

Table 2.3: Genes involved in human atrial fibrillation and related arrhythmias identified by GWAS. Gene symbol, gene type and heart function involved are shown.

Gene symbol	Gene type	Heart function involved
Tbx5	T-box Transcription factor	Atrial fibrillation; PR interval
Tbx20	T-box Transcription factor	QRS duration
Cav1	Caveolae protein	Atrial fibrillation; PR interval
Zfhx3	Zinc finger homeobox protein	Atrial fibrillation
Cacna1d	Calcium channel voltage dependent	Atrial fibrillation; QRS duration
Kcnq1	Potassium voltage-gated channel	Atrial fibrillation; QT interval

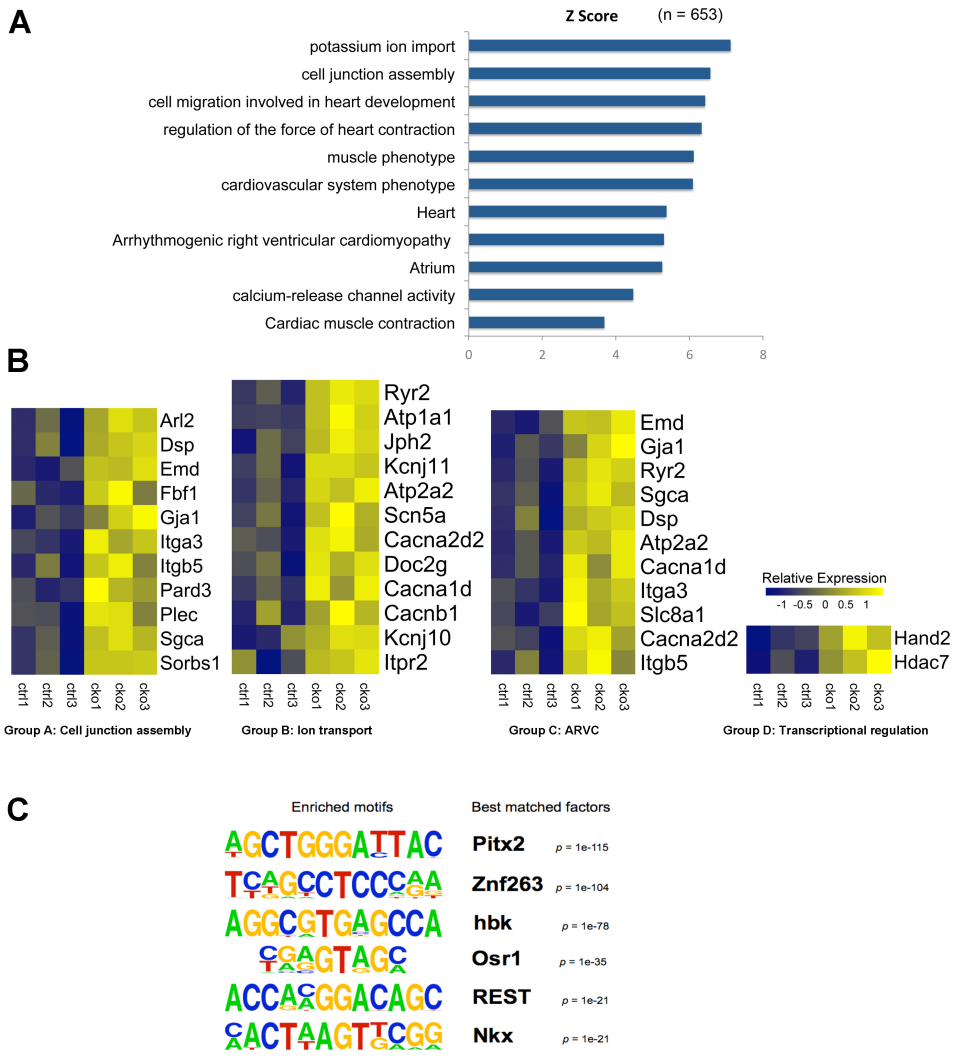


Figure 2.6: Overlay of Pitx2 ChIP-Seq and gene expression profiling assays for *Pitx2* CKO heart. (A) Gene ontology (GO) analysis for genes (*n*=653) from overlay of Pitx2 ChIP-Seq and Pitx2 CKO microarray assays (genes in close proximity to Pitx2 binding chromosomal regions identified by Pitx2 ChIP-Seq assay that also show significant change in microarray assay for *Pitx2* CKO heart). Both assays performed on 3-month-old adult mouse hearts. (B) Heat map of gene expression profiling obtained by *Pitx2* CKO microarray assay. The color bar represents relative expression level of log-transformed value for each gene across different samples. Ctrl1-Ctrl3: control mouse heart (*Pitx2* $^{Flox/Flox}$); CKO1-CKO3: *Pitx2* CKO mouse heart (*Pitx2* $^{Flox/Flox}$ *MCK-Cre*). (C) Motif analysis for peaks from Pitx2 ChIP-Seq (*n*=11,280). Enriched DNA binding motifs and best matched factors are shown.

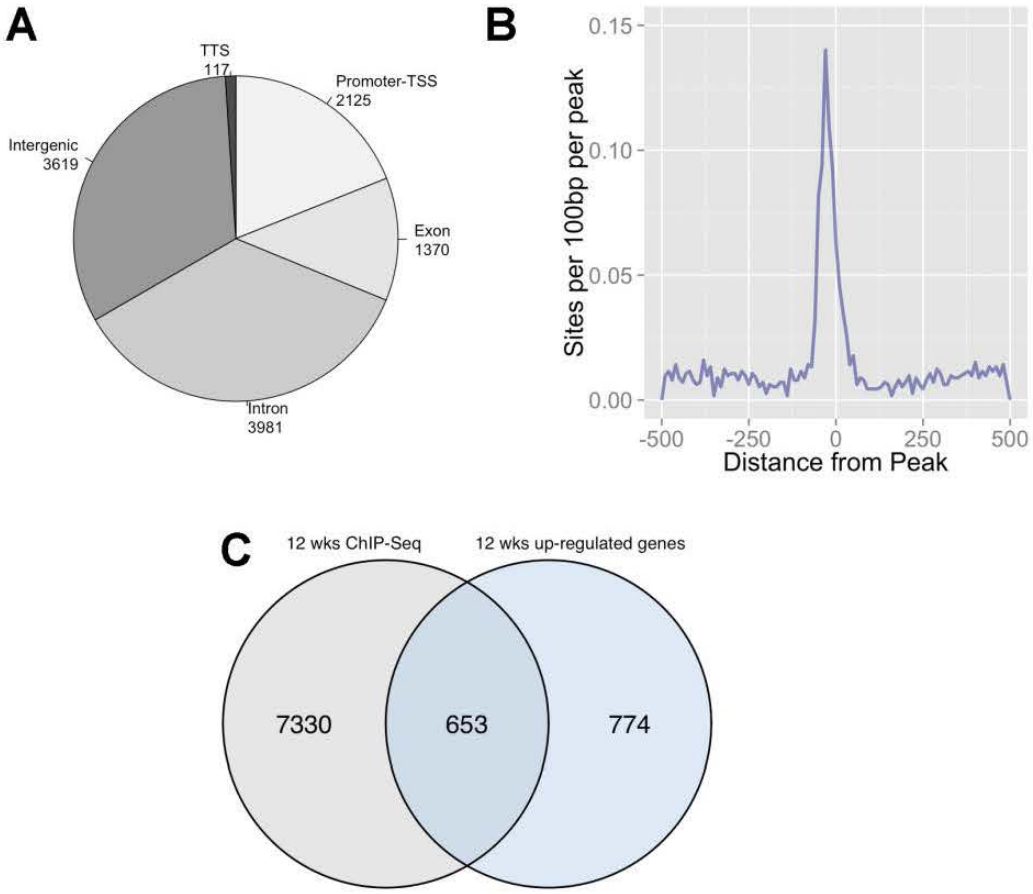


Figure 2.7: ChIP-Seq peaks overlay with upregulated genes from microarray. (A) Annotation of Pitx2 ChIP-Seq peaks from 12-week-old heart. Total 11,275 peaks were detected. 18.8% of the Pitx2 ChIP-Seq peaks were located in promoter or transcription start sites (TSS) regions; 12.1% in exons; 35.3% in intron, 32.0% in intergenic regions and 0.1% in transcription terminal sites (TTS). (B) The density of Pitx2 motif (GCTGGGATTACA) within ChIP-Seq peaks was shown. The occurrence of Pitx2 motif was close to the center of the peaks. (C) Total 7,983 unique genes were found by annotating closest gene to the ChIP-Seq peaks. 1,427 genes were up-regulated in 12-week-old Pitx2 mutant hearts. Overlaid 653 genes consist of 8.1% genes having ChIP-Seq peaks and 45.7% up-regulated genes in 12-week-old Pitx2 mutants.

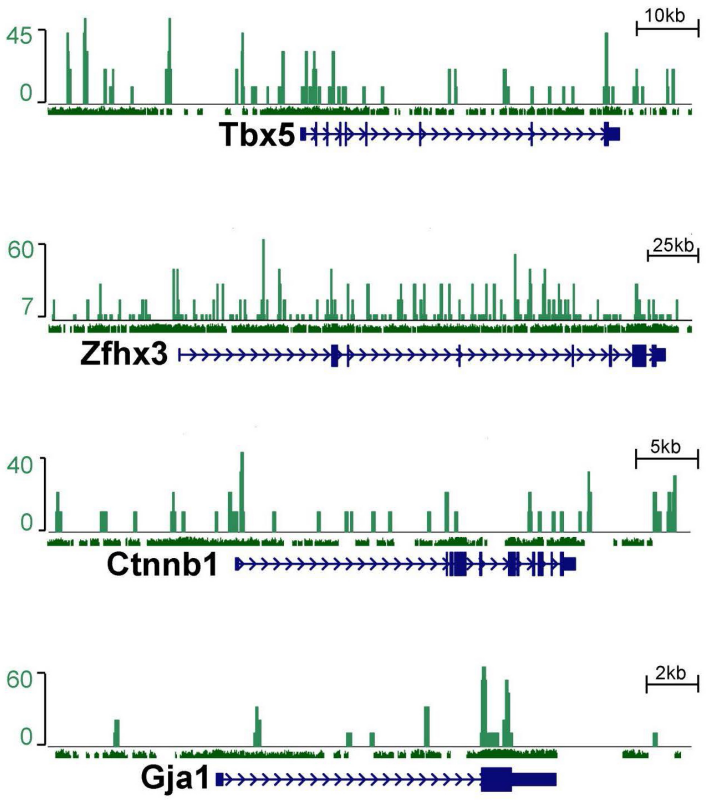


Figure 2.8: Genome browser tracks for additional potential targets of Pitx2 in adult heart identified by ChIP-Seq. Four potential targets of Pitx2 identified by ChIP-Seq by our studies are shown here: *Tbx5*, *Zfhx3*, *Ctnnb1* and *Gja1*. Peaks from ChIP-Seq and conservation with human genome are shown. Normalized ChIP-Seq tag numbers are shown on the *y*-axis.

normal pituitary development as is Pitx2 [104]. Caveolins are known to associate with channel proteins and are thought to modify channel function [144]. Our results also revealed higher mRNA expression levels for the genes *Gja1*, *Emd*, *Dsp*, and *Plec*, encoding ID components that are required for structural homeostasis and signaling between cardiomyocytes (Figure 2.9B). Genes that were unchanged or reduced in reverse transcription PCR assays are listed in Table 2.4.

Based on phylogenetic conservation and Dnase I hypersensitivity, we cloned 1 to 3 kb DNA sequences flanking multiple Pitx2 ChIP-Seq peaks (Figure 2.9A) into luciferase reporters and performed transactivation experiments in cultured cells. Luciferase activity for the ChIP-Seq regions that are in close proximity to *Tbx20*, *Cacna1d*, *Kcnq1*, *Cav1*, and *Emd* were repressed significantly on Pitx2 co-transfection, indicating that Pitx2 directly represses gene expression through binding to these chromosomal regions (Figure 2.9C).

2.3.5 Pitx2 mutants have an atrial cardiomyopathy with disrupted junctional complexes

Among the top Pitx2 target gene function categories are cell junction assembly and arrhythmogenic right ventricular cardiomyopathy. To investigate ID status in *Pitx2* CKO hearts, we used immunostaining with antibodies against β-catenin to label adherens junctions in the ID. β-catenin mediates cell adhesion between cardiomyocytes and also functions as a cell signaling molecule once activated by Wnt signaling. In controls, β-catenin clearly marked IDs, whereas in *Pitx2* CKO, there were fewer distinct IDs and more β-catenin expression on lateral aspects of cardiomyocytes (Figure 2.10A-D).

We used TEM to investigate further ID structure in *Pitx2* mutants. In contrast to controls (Figure 2.10E-G), *Pitx2* mutant IDs had widened spaces between junctions,

Table 2.4: Expression unchanged or reduced gene tested by qRT-PCR. Other gene tested by qRT-PCR. Gene symbols and result from the experiments are shown

Gene symbols	Alteration in Pitx2 CKO left atrium
Kcnn3 Nkx2.5 Ctnnb1 Tbx5 Dlg1	No significant change
Cav2	Decreased

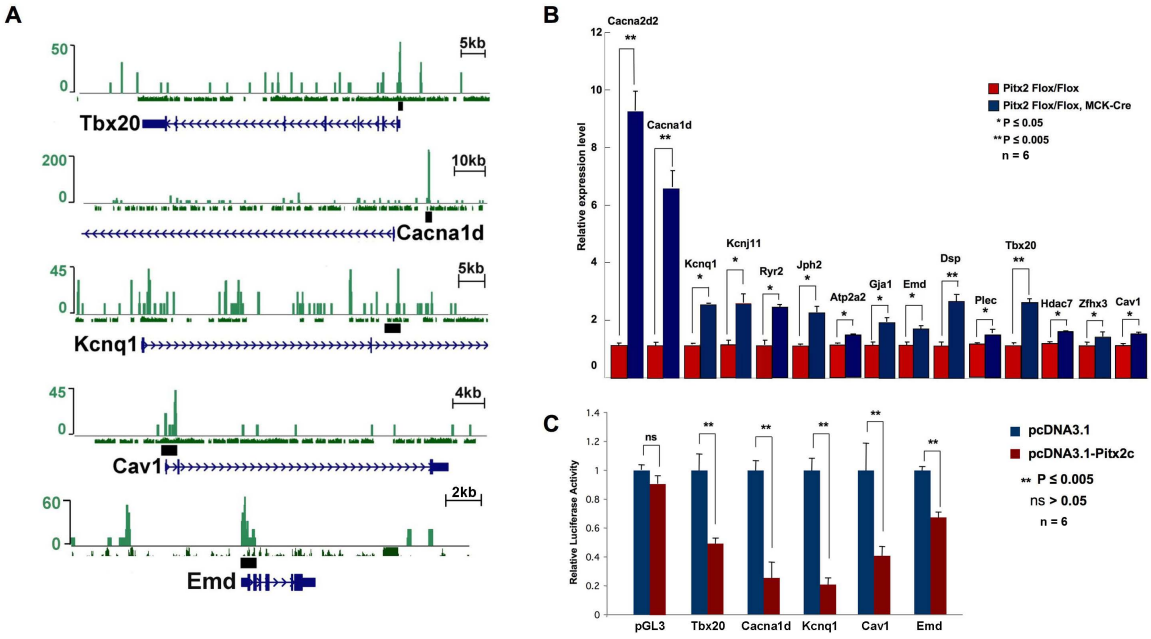


Figure 2.9: Validation of Pitx2 targets from ChIP-Seq and microarray assays. (A) Genome browser tracks for potential targets of Pitx2 in adult heart identified by ChIP-Seq. Peaks from ChIP-Seq, conservation with human genome and the chromosomal regions used in reporter assay are shown. Normalized ChIP-Seq tag numbers are shown on the *y*-axis. (B) Alteration in expression levels for the potential Pitx2 targets were detected by realtime RT-PCR in the left atrium of 3 to 4-month-old control and *Pitx2* conditional knock out mice. (*n*=6) (C) Reporter assay for ChIP-Seq identified potential targets of Pitx2. Chromosomal regions were named after the genes in closest proximity. Values and error bars represent mean and standard deviation (*n*=6). Figure credit: Lele Li, Ye Tao

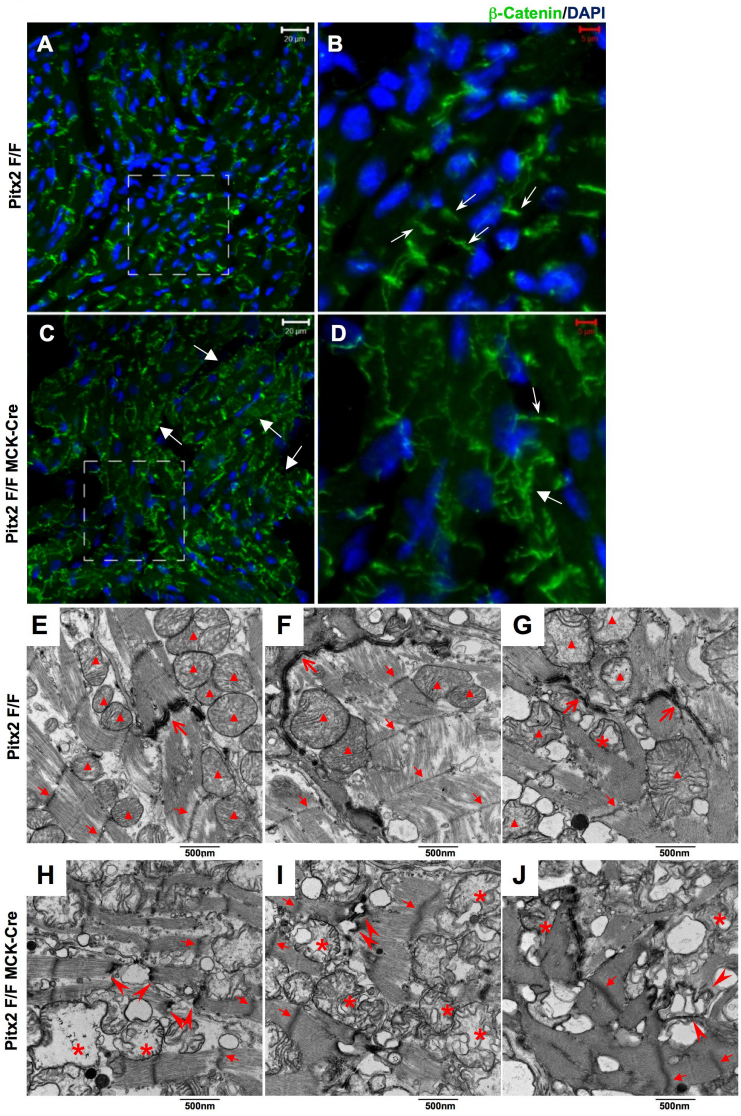


Figure 2.10: *Pitx2* CKO mouse heart has altered expression pattern for β-catenin and disrupted intercalated discs (IDs). (A-D) β-catenin immunostaining on left atrium sections of 3-month-old control (*Pitx2* $^{Flox/Flox}$) (A-B) and *Pitx2* CKO (C-D) mice shows enhanced expression of β-catenin on cardiomyocytes lateral aspect (big arrows in D) and in cytoplasm (big arrows in C). Small arrows point to IDs. (E-J) Electron microscopy shows disrupted IDs in one-year-old *Pitx2* CKO heart. (E-G) in control left atrium, IDs (big arrows) and mitochondia were intact (triangles). (H-J) in *Pitx2* CKO left atrium, some IDs were disrupted (arrowheads). Large number of mitochondria were deteriorated (asterisks). Indistinct Z-lines were observed in *Pitx2* CKO left atrium (small arrows). Figure credit: Ye Tao

indicating ID deterioration (Figure 2.10H-J). We also found evidence for mitochondrial dysfunction with swollen and vacuolated mitochondria in *Pitx2* CKO hearts (Figure 2.10H-J). *Pitx2* CKO cardiomyocytes also had indistinct Z-lines, indicating that Pitx2 is required for maintaining normal sarcomere cytoarchitecture (Figure 2.10H-J).

2.4 Discussion

Pitx2 is the gene in proximity to the 4q25 familial AF locus. Previous work indicated that reduced Pitx2 levels result in AF predisposition. Before our experiments, it was unknown whether *Pitx2* was important in postnatal heart because previously reported adult phenotypes could result from developmental defects. In addition to uncovering a *Pitx2* postnatal function, our findings reveal novel Pitx2 target genes and that Pitx2 is required for ID homeostasis in postnatal atrium (Figure 2.11).

2.4.1 Pitx2 postnatal deletion results in sinus node dysfunction

Pitx2 deletion in postnatal atrium caused resting atrial arrhythmias, indicating that in addition to its developmental function, Pitx2 has a genetically separable postnatal function in left atrial homeostasis. The predominant arrhythmia we observe with resting telemetry is SND, which is known to be clinically linked to AF in human patients. SND can result from defective impulse generation from the sinus node. In addition, SND similar to AF is associated with diffuse atrial remodeling, resulting in an arrhythmogenic substrate. Consistent with this, we observe anatomic disruption of cardiomyocyte junctions in Pitx2 CKO hearts [69].

The *Pitx2c*$^{+/-}$ mice that have a 37% reduction in Pitx2c levels have no structural abnormalities or obvious histological defects but are still prone to AF [63]. Molecular analysis indicates that *Pitx2c*$^{+/-}$ mice have electric remodeling with changes in channel gene expression. The authors also noted changes in desmosomal genes, such as *Dsg2*, raising the possibility that *Pitx2c*$^{+/-}$ mice may also be prone to structural remod-

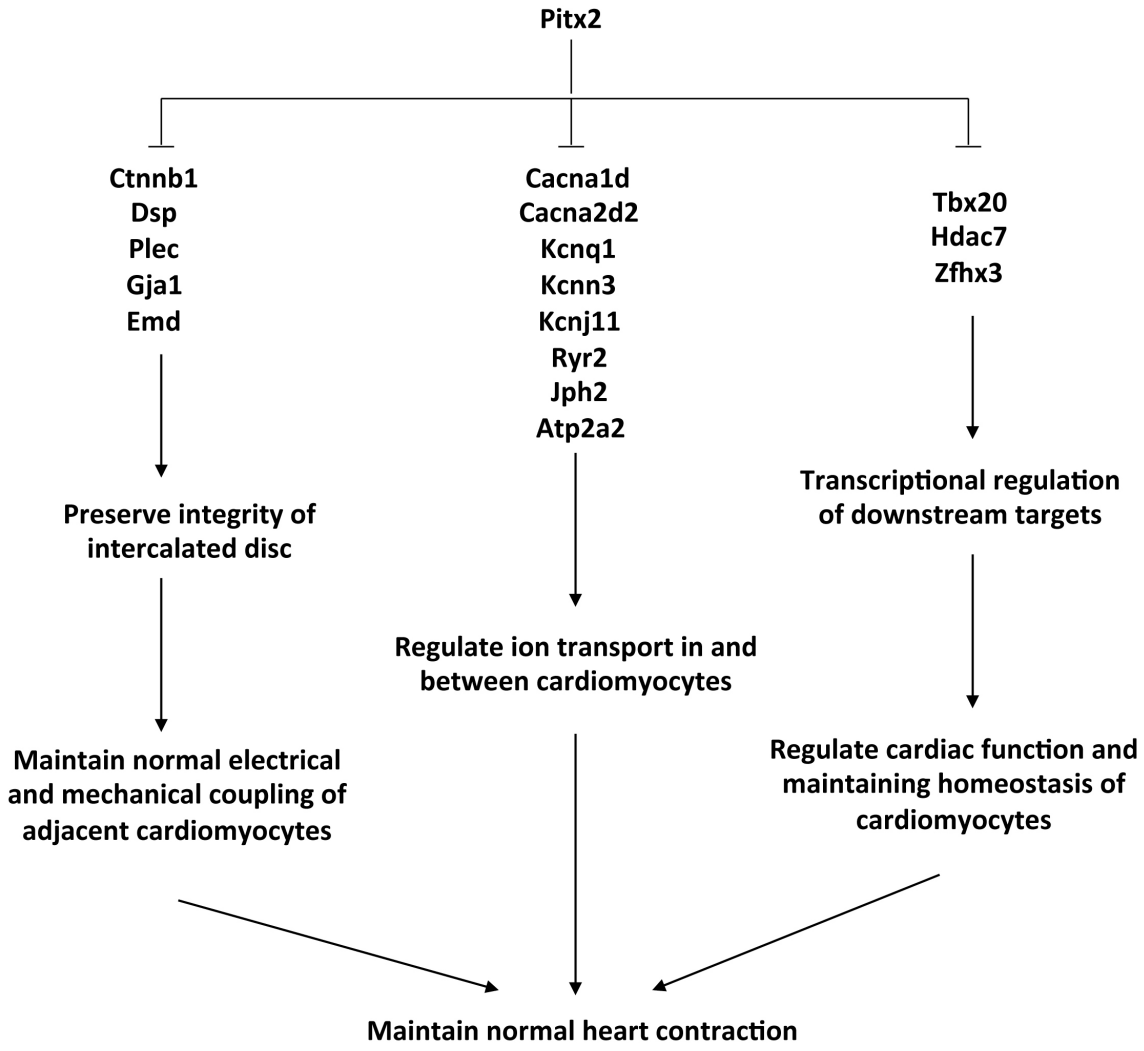


Figure 2.11: Model of Pitx2's function in postnatal heart. In postnatal heart, Pitx2 binds to its targets through DNA binding domain and regulates expression of targets. Three groups of genes were identified to be targets of Pitx2. The first group contains genes regulating cell adhesion/cell junction assembly, which is critical for preserving normal structure of ID. The second group contains genes that regulate ion transportation in and between cardiomyocytes, which are important for normal conduction of electrical signal that induces heart contraction. The third group contains transcription regulators that regulate a large number of downstream targets and involved in regulating cardiac function and maintaining homeostasis of cardiomyocytes.

eling. Taken together with previous studies, our findings reveal that *Pitx2* has both developmental and postnatal functions that predispose the heart to atrial arrhythmias.

2.4.2 Pitx2 regulates intercalated disc maturation and homeostasis

Our array data reveal that major changes in gene expression were detected at 12 weeks of age. Moreover, ChIP-Seq data support the model that Pitx2 directly regulates genes involved in ID homeostasis. Structural remodeling of intercellular connections and an increase in fibrosis represent known anatomic substrates for AF. Notably, *Pitx2* mutants have upregulated expression of ID genes, suggesting that the correct stoichiometry of ID component genes is important to maintain ID structure. The immunofluorescence showing β-catenin mis-localization and TEM revealing breakdown of intercellular junctions also strongly support the notion that *Pitx2* is required for ID maturation and homeostasis.

In the developing and postnatal heart, ID components including desmosomes, gap junctions, and adherens junctions are initially expressed in a dispersed pattern. As the heart matures, ID components become localized into the ID at the cardiomyocyte termini [5]. At the transcriptional level, expression of genes such as Emd and Gja1 is reduced as the heart matures [136, 119]. Intriguingly, emerin (Emd) is known to function as a β-catenin-interacting protein that also modulates β-catenin intracellular localization [143]. It is conceivable that lateralized β-catenin that we observe in the *Pitx2* mutant left atrium may be because of elevated Emd levels. Together, our data indicate that *Pitx2* is required for this transcriptional downregulation of ID component genes in left atrium myocardium.

Our findings have implications for AF progression to chronic AF. The severe anatomic abnormalities observed in *Pitx2* CKO hearts suggest that these are irreversible structural defects. Moreover, the TEM data suggest that the *Pitx2* CKO hearts have dete-

riorating mitochondria, suggesting cellular compromise further supporting the notion that Pitx2 CKO arrhythmias are irreversible. The TEM data also revealed that *Pitx2* CKO atrial cardiomyocytes had Z-line defects, suggesting that Pitx2 also regulates sarcomere homeostasis. Interestingly, we previously observed formation of M-lines and more mature sarcomeric structure in *Pitx2* CKO extraocular muscle, which normally lacks M-lines, suggesting context-dependent functions for Pitx2 in sarcomere regulation [157].

2.4.3 Pitx2 and ion current regulation

Our data support previous observations that multiple genes involved in calcium handling and potassium channels are changed in *Pitx2* mutants [63]. The ChIP-Seq data indicate that multiple genes that compose the calcium release unit such as *Ryr2*, *Jph2*, and *Atp2a2* are regulated directly by Pitx2 [28]. Genes involved in calcium homeostasis have been implicated in both AF initiation and progression to chronic AF [122]. For example, Ca^{2+}/calmodulin-dependent protein kinase II has been shown to phosphorylate the ryanodine receptor 2 at S2814, resulting in incomplete ryanodine receptor 2 closure with calcium leak. Increased phosphorylation of ryanodine receptor 2 has been demonstrated in human patients with chronic AF, as well as other AF models [15, 29]. Our ChIP-Seq/microarray data sets indicate that RyR2 is upregulated in *Pitx2* mutant adult hearts, suggesting that in addition to phosphorylation, transcriptional regulation may be a regulatory node for modulating calcium handling in atrial cardiomyocytes.

In *Pitx2* CKO postnatal hearts, the inward rectifier *Kcnj11* was both directly bound by Pitx2 and upregulated at the mRNA level, a likely AF risk factor [60]. This upregulation is consistent with the shortened action potential duration that has been described previously in *Pitx2* mutants [19, 63].

2.4.4 Pitx2 and AF in the developing and postnatal heart

Our findings suggest that postnatal atrial cardiomyocytes require persistent Pitx2 expression for ID homeostasis and maturation. Pitx2 regulates separate mechanisms, a developmental mechanism and a postnatal homeostatic mechanism, that result in an arrhythmogenic substrate. This is also reflected in the expression dynamics of Pitx2 target genes in the developmental and postnatal contexts. *Scn5a* and *Kcnj10* are downregulated in *Pitx2* mutant atrium during development but upregulated in postnatal mutants [19]. One explanation for this is that Pitx2 may have separate cofactors at developmental and postnatal stages that result in distinct transcriptional readouts.

Our findings also have implications for human AF. Previous GWAS implicated multiple genes as potential contributors to the pathogenesis of human AF and prolonged PR interval [84]. Our Pitx2 ChIP-Seq data identified several of these genes as potential Pitx2 target genes in postnatal hearts (Table 2.2 [84, 100]. Because Pitx2 regulates many of these genes in the postnatal and adult heart, it is conceivable that drugs can be developed to modulate the molecular interaction between Pitx2 and its target genes. Because fetal therapies currently lack feasibility, our findings that many important Pitx2-regulated events occur postnatally strengthens the likelihood that Pitx2-mediated AF may be treatable in the future.

2.5 Materials and methods

2.5.1 Mouse alleles and transgenic lines

The *MCK-Cre* transgenic line, *R26R* transgenic line, *Pitx2* conditional null or floxed allele and Flag knock in allele have been described [138, 156]. The *Pitx2* conditional null allele contains LoxP sites flanking the exon encoding the N-terminus of homeodomain and conditionally removes function of all *Pitx2* isoforms.

2.5.2 Telemetry ECG

Implantation of the telemetry transmitter ETA-F10 (Data Sciences International) was done following the protocol of the manufacturer. Transmitters were placed intraperitoneally and ECG leads were at a lead II configuration. ECG data was collected by DSI Telemetric Physiological Monitor System and processed by Dataquest A.R.T. 3.1 software (Data Sciences International) without stress from restraint or anesthesia.

2.5.3 ChIP-Sequencing

Chromatin immunoprecipitation was performed as described previously using EZ ChIPTM kit (Millipore) and Anti-FLAG M2 Affinity Gel (Sigma) . For library construction, 10-50 ng of purified ChIP DNA was end-repaired using Ion Plus Fragment Library Kit (Part no. 4471252, Life Technologies) and purified with two rounds of AMPure XP beads (Beckman Coulter, Brea, CA) capture to size select fragments 100-250 bp in length. End-repaired DNA was ligated to Ion Torrent compatible adapters. Followed by 15 cycles of PCR amplification before downstream template preparation. DNA fragments of completed library ranged from approximately 170-220 bp in length. Sequencing was performed on Ion Torrent PGM system (Life Technologies).

2.5.4 Statistical analysis

Microarray raw data were generated using Affymetrix MAS5.0, global scaling was performed to set mean target intensity to 100. The normality of the microarray data was checked using quantile-quantile plot and Shapiro-Wilk test on R (3.0.1). Differential expressed genes were detected using R package *limma* (3.16.8) [121] with threshold p-value ≤ 0.05 and fold change ≥ 1.5. Benjamini-Hochberg correction was used to calculate false discovery rate. All the differential expressed genes have FDR < 0.35. Clusters were automatically detected using R package *HOPACH* (2.20.0) [101]. The

distance matrix was calculated using cosine angle method. Over-represented gene ontology terms were identified using GO-Elite [152] with Z score > 3, FDR < 0.10. FDR was calculated using Benjamini-Hochberg correction after permutation (n = 5000). qRT-PCR and luciferase reporter assay results were tested using two-sample Wilcoxon test in R. Difference of R-R interval between wild type and mutant mice was tested using two-tailed, unpaired Student's t-test using R ($n > 400$). ChIP-Seq peaks were identified using Homer [49] under the assumption that the local density of the reads follow an Poisson distribution. The peaks were called using cutoff: read number enrichment ≥ 4 folds, FDR < 0.001, FDR effective Poisson p-value < 3.7e-8, and minimum read number 5.

2.5.5 β-Gal staining and immunostaining

β-Gal staining was done as previously described [138]. Immunostainings were done on frozen sections of adult heart. Samples were incubated in primary antibody at 4°C overnight. After washing in PBS, sections were incubated in the appropriate fluorescent-labeled secondary antibodies, followed by counterstaining with DAPI (Sigma) then mounted in Vectashield H-1000 mouting medium (Vector Laborataries, Inc.). Primary antibodies used were as follows: Rabbit polyclonal anti-Pitx2 (1:100) (Capra Science) and Rabbit anti-β-catenin (1:100) (Cell Signaling). Secondary antibodies used was as follows: Alexa Fluor 488 goat anti-rabbit IgG (1:1000) (Molecular Probes). Immunofluorescent images were captured on a Leica TCS SP5 confocal microscope, and all functions were controlled via Leica LAS AF software (Leica Microsystems).

2.5.6 Realtime RT-PCR and luciferase assay

Total RNA was isolated using the miRNeasy mini Kit (QIAGEN), followed by RT-PCR using qScriptTM cDNA SuperMix (Quanta Biosciences). Three biological repli-

cates were collected for control and mutant. Real-time thermal cycling was performed using Stepone Plus thermal cycler (Applied Biosystems) with SYBR Green (Quanta Biosciences). Comparative Ct method was used to quantitate relative gene expressions. Firstly, Ct value of both mutants and controls was normalized to an endogenous housekeeping gene (GAPDH). Secondly, each sample was calibrated to the mean of control Ct value. Lastly, Ct value was transformed to fold change using exponential transformation. Standard deviation of both control and mutant samples were calculated within the groups. Two-sample Wilcoxon test was performed between control and mutant. Sequences of PCR primers are available upon request. Luciferase reporters were constructed from PGL3 luciferase reporter vector (Promega) and 1-3 kb DNA fragment amplified from Genomic DNA using primers for ChIP-Seq peak regions. PCR was performed using PrimeSTAR HS DNA Polymerase (Takara Biotechnology Co., LTD.). Clonings were done by using Gateway Vector Conversion System (Invitrogen). Sequences for PCR primers are available upon request. Expression contruct for Pitx2c was obtained from Dr. Brad Amendt. 293FT cells were transfected with the reporter construct and Pitx2c expression construct or control construct.Transfection was done by using Lipofectamine 2000 (Invitrogene) per the manufacturer's protocol. Luciferase activities were measured on Infinite M200 Pro multimode microplate reader (TECAN) using Dual-Glo Luciferase Assay System (Promega) per the instruction of the manufacturer. Average luciferase activity are reported with standard deviation. Two-sample Wilcoxon-test was performed using R.

2.5.7 R-R interval measurement

Telemetry ECG data was collected on DSI Telemetric Physiological Monitor System and processed by Dataquest A.R.T. 3.1 software (Data Sciences International). For each mouse, manually select uninterrupted one miniute ECG interval with uni-

form maximum R-wave voltage; number of heart beats counted accordingly. R-R intervals were measured in the unit of millisecond. Mean and standard deviation of R-R interval from each mouse were calculated using R software. Two-tailed, unpaired Student's t-test ($n > 400$) was performed for mean and standard deviation values of R-R interval; no significant difference was detected of mean R-R interval value between controls and mutants ($p = 0.54$); standard deviation of R-R interval from mutant mice were significantly greater than that of control mice ($p = 0.003$).

2.5.8 Microarray and ChIP-Seq data processing

Pitx2 control and mutant hearts were collected from 3-, 6- and 12-week-old mice. At each time point, three controls and three mutants were collected as biological replicates. cDNA microarray analysis was performed using Affymetrix GeneChip Mouse Genome 430 2.0 Array (Affymetrix, Santa Clara, CA). The raw data of microarray was generated using Affymetrix Microarray Suite version 5.0 (MAS 5.0) using Affymetrix default analysis settings and global scaling as a normalization method. The trimmed mean target intensity of each array was arbitrarily set to 100. The value distribution of raw intensity from total 18 samples was graphically viewed in box plots. To show reproducibility across three biological replicates within each group, linear correlation coefficient value r was measured between each pair of replicates within each groups using R software function *var*.

Differential expressed genes were detected using R bioconductor package *Limma* version 3.16.8. Briefly, log transformation was performed to the raw intensity values, the normality was checked using histogram and quantile-quantile plot for each condition; Shapiro-Wilk test was performed on randomly selected 300 genes from each array using R shapiro test function, resulted p value > 0.05 for all 18 individual chips. Student's t-test was performed to identify differences of expression level between wild

type and mutants for each probe. Benjamini & Hochberg false discovery rate method was used for multiple comparsion correction. Differentially expressed genes were called using cutoffs p-value ≤ 0.05; fold change ≥ 1.5 and FDR ≤ 0.35.

We compared differential expressed genes in 3-, 6- and 12-week of age individually. Total 3,323 genes were collectively called as differentially expressed, and were subjected to R package *HOPACH* to perform unsupervised clustering based on a cosine angle distance matrix. The first level of the hierarchical tree contains 5 clusters, one of which has 2,697 genes with significant higher expression value compared to the other 4 clusters. Gene ontology analysis showed that genes from this cluster are enriched in heart and atrium. Clusters I-V, shown in Figure 2.3, were clusters on the third level of the tree derived from this highly expressed, heart-related gene group, with a minimum of 30 genes in each cluster. Raw intensity values from individual samples were used to plot in heat map. Relative expression value was calculated using following formula and illustrated in color:

$$\textit{Relative expression value} = \frac{\textit{Individual expression density} - \textit{mean}}{\textit{standard deviation}} \tag{2.1}$$

Over-represented gene ontology terms were identified using GO Elite with its default Z score and permutation methods. To calculate Z score, the observed number of genes in the input gene list was subtracted from the expected number of genes associated with a particular biological term and divided by the standard deviation of the observed number. To determine the likelihood of observing the calculated Z score by chance, the permutation analysis was performed. The same number of genes as input gene list was randomly selected from background gene list and calculated Z score. This process was repeated 5000 times for each input gene list. False discovery rate

was calculated using Benjamini-Hochberg correction. GO terms were plotted using *Z* score (Figure 2.6A). Only terms with FDR < 0.10 were reported.

Ion torrent PGM reads were mapped to the mm9 assembly (NCBI Build 37) using Torrent Suite (2.0.1) aligner Tmap (0.2.3) (Life Technologies). The ChIP-Seq signal was normalized to a 10 million reads total and visualized in the UCSC genome browser. ChIP enriched peaks were identified using Homer 4 by the default setting. Briefly, a fixed peak size was calculated automatically, in our case, the peak size was 162 bp. The program then scanned the entire genome for 162 bp clusters with the highest density of reads. After all clusters have been found, the program compared Pitx2 ChIP-Seq reads against input control reads. The cutoffs for calling the peaks were read number enriched at least 4 fold, FDR threshold 0.001, FDR effective Poisson *p*-value 3.7e-8, and minimum read number 5. This process produced 11,280 enriched peaks. Nearest genes associated to the enriched peaks were annotated using the annotatePeaks function and de novo motif discovery was performed using the *findMotifsGenome* function in Homer. Genes associated to ChIP-Seq peaks were overlaid with significantly up-regulated genes in 12-week mutants. The overlaid gene list was analyzed with GO Elite for enriched gene ontology terms. Microarray and ChIP-Seq data are available through the NCBI Gene Expression Omnibus (GEO) data repository under accession GSE50401.

3. LONG-RANGE INTERACTIONS BETWEEN *PITX2* AND HIGH-RISK LOCI OF ATRIAL FIBRILLATION

Human 4q25 region is highly associated with risk of familial atrial fibrillation (AF), with multiple single nucleotide polymorphisms (SNPs) at this locus. The region was first identified from a genome-wide association study (GWAS) performed in an Icelandic population [41]. Multiple subsequent studies have consistently demonstrated a robust association between this locus and AF [84]. The AF-associated SNPs at the 4q25 locus are located in an gene-desert sized about 1.5 million base pairs. The closest gene is the paired-like homeodomain transcription factor 2 (*PITX2*). Previous work indicates haploinsufficiency of *Pitx2* promotes an arrhythmogenic phenotype in mice[138]. Moreover, as described in the previous section, *Pitx2* has its independent role in postnatal stage, regulating genes encoding cell junction proteins, ion channels, and critical transcriptional regulators. Importantly, many of these *Pitx2* targets have been implicated in human AF. However, the direct connection between non-coding 4q25 locus and *PITX2*is unclear. It is unknown whether the high-risk loci are regulatory elements that directly control the expression of Pitx2 via physical interaction; whether these regulatory elements alters developmental programs or triggered by environmental factors during adulthood, and whether up-stream signals, such as arrhythmia-induced cardiac stress, regulate Pitx2 expression through modifying topology of three-dimensional (3D) chromosome structure.

3D chromosome structure of the genome is known to play an important role in transcriptional control of genes for decades [22]. But not until recently, development of chromosome conformation capture (3C) technology [25] advanced our understanding of regulatory interactions in a three-dimensional setting. Coupled with

high-throughput sequencing, the technology now allows the assessment of one-to-all (4C), many-to-many (5C) or all-to-all (Hi-C) [23, 24] interactions. Moreover, methods such as ChIA-PET allows assessment of all loci bound by a specific protein ([36]). High resolution Hi-C results reveal the genome-wide organization of topologically associated domains (TADs) [73, 108]. TADs [27, 95, 117, 23] are defined as chromosomal regions that sequences are preferentially contact each other. The median size of TADs is 185 kb, but varies from 40 kb to 3Mb. The boundaries between TADs are highly associated with sites bound by CCCTC-binding factor (CTCF) [108], as well as cohesin complex. Here, we describe 4q25 is located in the same TAD with *Pitx2* which is located close to the boundary of the TAD. Bioinformatic analysis identified heart-specific enhancer activity overlaid with AF risk loci. 4C-Seq result suggest a direct interaction between the enhancer peaks and *Pitx2ab*and *Pitx2c*promoters, as well as a downstream gene encoding a glutamyl aminopeptidase (*Enpep*). Knocking-out these enhancers using CRISPR/Cas9, *Pitx2ab* expression significantly reduced *in vivo*. Furthermore, deletion of CTCF binding sites reduce the expression of *Pitx2ab*, *Pitx2c* and *Enpep*. Our findings, revealing a direct interaction between *Pitx2* and AF high-risk locus, unveils a potential regulatory functions of non-coding variations in atrial fibrillation.

3.1 Results

3.1.1 Pitx2 is located on the topologically associated domain boundary

Human 4q25 AF associated gene desert region is highly conserved with mouse a 1 Mb block in mouse chromosome 3, both in DNA sequence and genome organization levels. Within this 1.5 Mb genome block, *Pitx2* and its neighboring genes are located by the same order and orientation. More over, high resolution Hi-C [108] and CTCF ChIP-Seq [33] datasets across different tissue revealed striking similarity

of TAD structure between human and mouse tissue (Figure 3.1), which most of then 4q25 gene desert region is located in one TAD, while *Pitx2* and its neighbor *Enpep*are located on the boundary of the TAD. CTCF binding sites are found in both *Pitx2* and *Enpep* gene bodies, as well as in-between these two genes. Interestingly, a CTCF binding peak separates two individual promoters of *Pitx2* isoforms, namely *Pitx2ab* promoter and *Pitx2c* promoter. ChIA-PET dataset of a cohesin complex protein [30], Sma1, revealed long-distance interactions between *Pitx2* promoters and regions in the same TAD, as well as local interaction between themselves (Figure 3.1A&D). Furthermore, indicated by histone marks ChIP-Seq datasets [33], *Pitx2* expression is active, whereas *Enpep*expression is poised in mouse ESC. Accordingly, interaction activity within *Pitx2* gene body is much higher than within *Enpep*[30], suggesting a correlation between transcription activity and chromatin organization.

3.1.2 Pitx2 up-stream regulatory region functions as an enhancer

SNPs that highly associated with AF are located 150 kb upstream of *PITX2*, span within a 100 kb window (Figure 3.2). To address whether this region has potential regulatory function, we aligned the DHS-Seq datasets across different tissues [110], and found that clusters of DHS activity in this region. Interestingly, two sets of these peaks are consistently observed in multiple fetal hearts, but not in other tissues, or in adult hearts (Figure 3.2A). Since DHS marks open chromosomal region in general, we aligned DHS genomic tracks with H3K4me1 ChIP-Seq tracks, which mark enhancer activity, and found that these heart-specific DHS peaks are overlaid with heart H3K4me1 peaks, indicating potential enhancer activity. Furthermore, using a wider range of markers, an integrative algorithm predicts the enhancer activity in this region [110]. Although these enhancer activity are limited in fetal hearts, another enhancer profiling dataset showed abnormal enhancer activity in this region from an adult donor

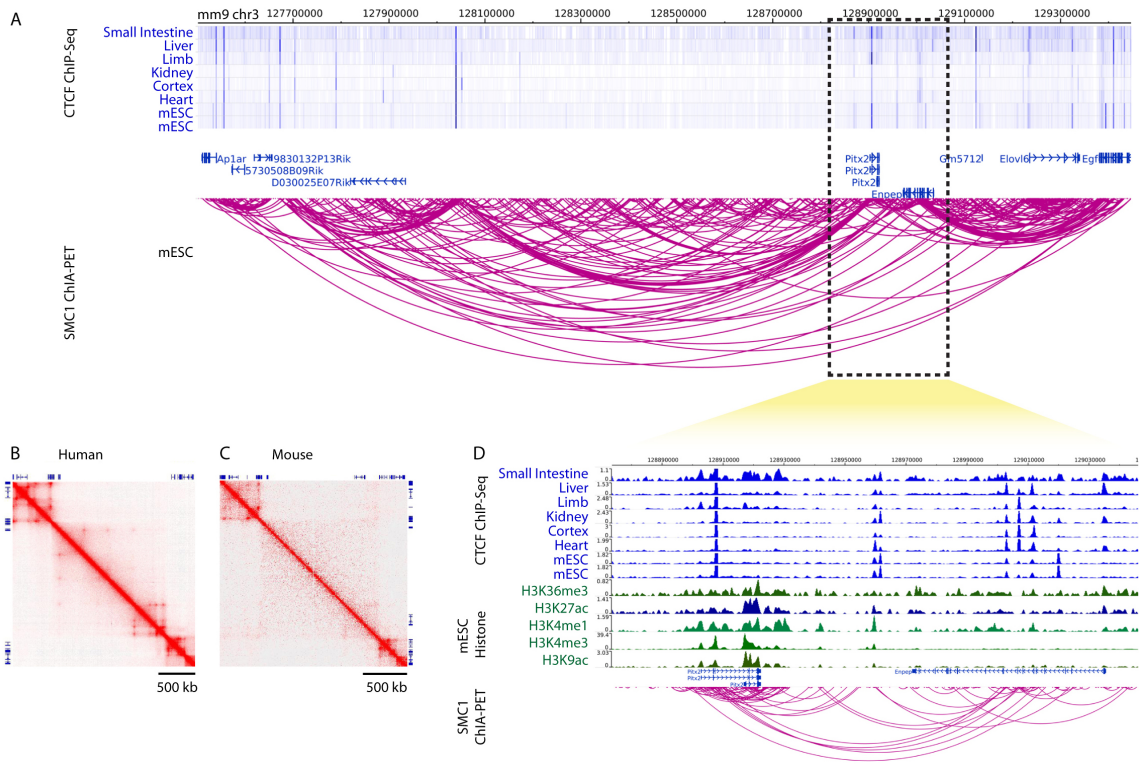


Figure 3.1: Conservation of topologically associated domains (TADs). (A) CTCF ChIP-Seq and Smc1 ChIA-PET in mouse tissue showed Pitx2 and Enpep are on the boundary between TADs. (B, C) TAD structure in human 4q25 region is conserved between human and mouse. (D) Local region of mouse *Pitx2* and *Enpep*. CTCF is associated with segregation between TADs. Additionally, chromatin activity is associated with chromatin interaction, which is indicated by Smc1 ChIA-PET.

with coronary artery disease (Figure 3.2A), suggesting this region can be potentially reactivated during disease state. Noticeably, individual SNPs do not exactly overlay with these enhancer peaks. Nonetheless, we pinpoint this 100 kb window contained multiple SNPs associated with AF risk is a heart-specific regulatory element that has enhancer activity.

To address whether we can study the function of this regulatory element in a mouse model, we compared DNA sequence conservation between human and mouse genome and found that regulatory elements that marked by DHS and enhancer marks are highly conserved despite that their surrounding non-coding sequences have poor alignment (Figure 3.2C). We located the mouse sequences corresponding to human AF-associated enhancers and found that the region is marked by H3K27ac in mouse heart across different developmental stages [96] (Figure 3.2D). Moreover, we performed motif enrichment analysis using enhancer DNA sequences from both human and mouse, and found consensus binding sites of Pitx2, Hnf4a, Nrf2, Ets and Nkx factors are significantly enriched (Figure 3.2E).

3.1.3 Direct interaction between enhancer and Pitx2 promoter

To address whether *Pitx2* promoter directly regulated by upstream enhancers through long-range interaction. We dissected postnatal day 8 (P8) mouse hearts and performed 4C-Seq. Sequences that physically interact with each other are ligated together and form circular DNA. To profile all the sequences that interact with our regions of interest, or view points, we used inverse PCR to amplified blind sequences that in the same circular DNA molecule with *Pitx2*ab promoter, *Pitx2*c promoter, *Enpep* promoter, and upstream enhancers (Figure 3.3A). The PCR products were then barcoded and subjected to sequencing on Ion Proton platform. We collected total 30 million reads from 10 different view points. For individual reads, we subtract sequences from viewpoints

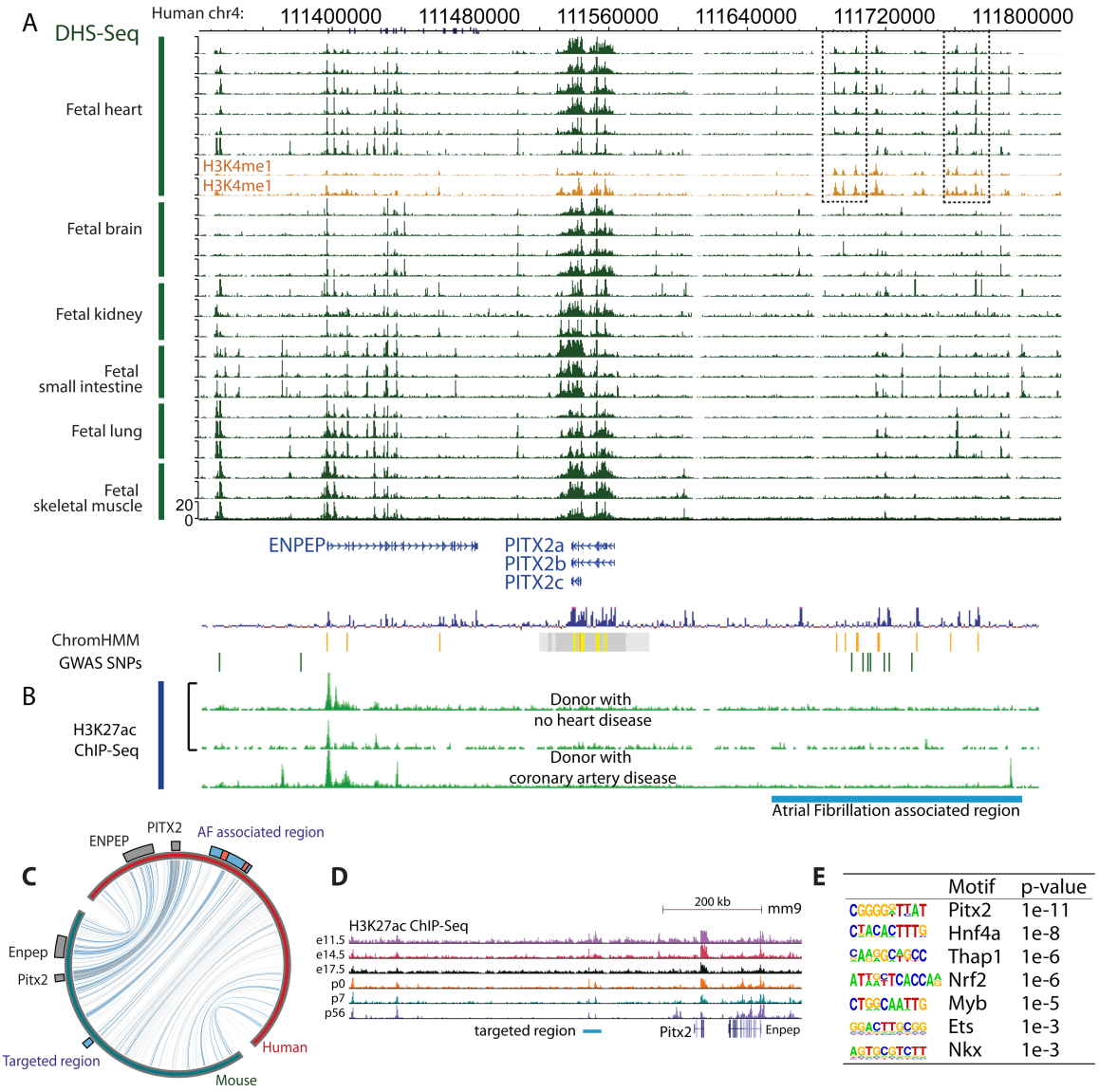


Figure 3.2: Atrial fibrillation associated region is overlaid with heart specific enhancers. (A) DnaseI Hypersensitivity Sites (DHS) shows human fetal heart specific region is located upstream of Pitx2. H3K4me1 ChIP-Seq and chromatin HMM analysis indicates enhancer activity in these regions. (B) Heart H3K27ac ChIP-Seq from donor with coronary artery disease show altered peaks in the upstream region of Pitx2 and downstream of Enpep. (C) Conservation alignments of DHS regions (light blue) and enhancers (red) between human and mouse genome. (D) Mouse heart H3K27ac ChIP-Seq tracks indicate conserved region of atrial fibrillation associated enhancers. (E) Enriched motifs present in the enhancer region.

and map them against mouse genome, and quantified reads that are ligated to each view point. Since each cell has only two copies of the chromatin, the amount of such read count reflects an averaged picture of contact frequencies between loci *in vivo*. Importantly, such read count is also a function of distance between loci [124]. Sequences that are proximal to each other are more likely to be cross-linked and ligated together. Using a linear regression model [37], we detected significantly enriched regions that directly contact with the view points. The AF-associated enhancer directly binds to *Pitx2ab*, *Pitx2c* and *Enpep* promoters, as well as the CTCF binding site in-between *Pitx2* and *Enpep* (Figure 3.3B). Furthermore, we repeat the 4C-Seq experiment in an embryo-derived teratocarcinoma cell line (P19), and found similar interaction existed.

To further investigate whether these AF-associated enhancers controls *Pitx2* expression. We deleted a 20 kb region that contains two H3K27ac peaks and conserved with human heart-specific enhancers via CRISPR/Cas9 system 3.3A). Two individual guide RNAs are designed two create double strand DNA breaks spontaneously and cause the deletion of large DNA fragments. We gained seven homozygous mutant cell lines and tested gene expression level of *Pitx2ab*, *Pitx2c* and *Enpep* using qPCR. Upon enhancer deletion, *Pitx2ab* expression is significantly reduced across all seven individual mutant lines by over 90%, whereas *Pitx2c* and *Enpep* expression do not show significant change in average (Figure 3.3C). However, the expression level of *Enpep* varies between different mutant lines (Figure 3.3C).

3.1.4 Enhancer-promoter interaction is mediated by CTCF

The deletion of upstream enhancers lead to significant reduction of *Pitx2ab* expression level, but not *Pitx2c*, suggesting a strong segregation between these two promoters. Known to be genome insulators, enriched CTCF binding sites are observed inside *Pitx2* gene body, separating two promoters of *Pitx2* isoforms. 4C-Seq results also sug-

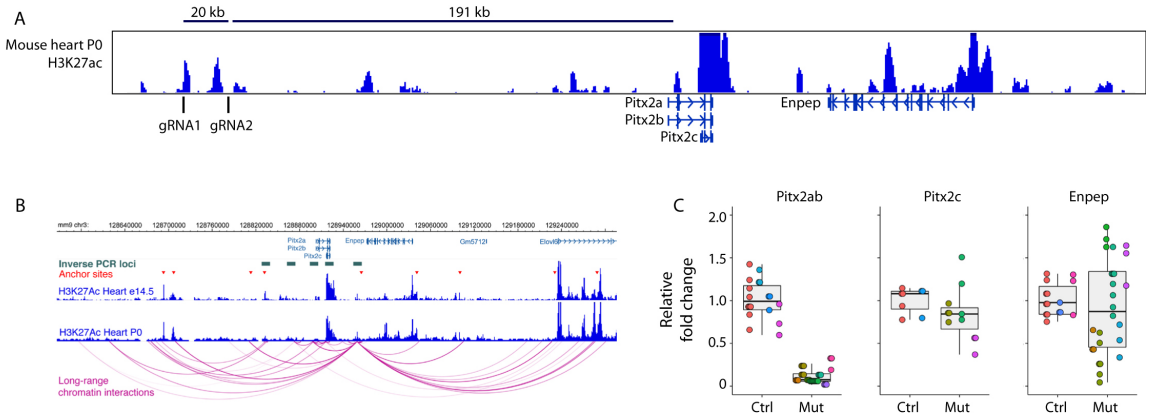


Figure 3.3: CRISPR mediated enhancer deletion. (A) Two guide RNAs are used to target 20 kb of enhancer region, which is located 191 kb upstream of *Pitx2ab* promoter. We obtained four females and one male mouse lines; and seven P19 homozygous cell lines. (B) High-throughput chromatin conformation sequencing (4C) indicates direct interaction between targeted enhancer and Pitx2ab, Pitx2c promoter and CTCF sites. (C) Individual gene expression of Pitx2ab, Pitx2c and Enpep are quantified in homozygous P19 cell lines using qPCR. Pitx2ab is significantly down regulated by the deletion of the enhancer. Pitx2c and Enpep show varied expression level among individual mutant cell lines.

gest direct interaction between *Pitx2* promoters and downstream CTCF binding sites. We therefore deleted six of these CTCF sites in P19 cells to address whether the regulatory function of *Pitx2* upstream enhancer is mediated by any of these CTCF sites (Figure 3.4A&B). We found that deletion of CTCF site in-between *Pitx2*ab and *Pitx2*c promoter (del 2, Figure 3.4A&B) increases *Pitx2*ab expression and decreases *Enpep* expression; whereas deletion of CTCF site downstream of *Enpep* leads to decreased gene expression of *Pitx2*ab, *Pitx2*c and *Enpep*. Furthermore, we deleted CTCF site that segregate *Pitx2*ab and *Pitx2*c promoters (del 2) in enhancer-deleted mutant cell lines, and found that the deletion of CTCF restored *Pitx2*ab expression level (Figure 3.4D).

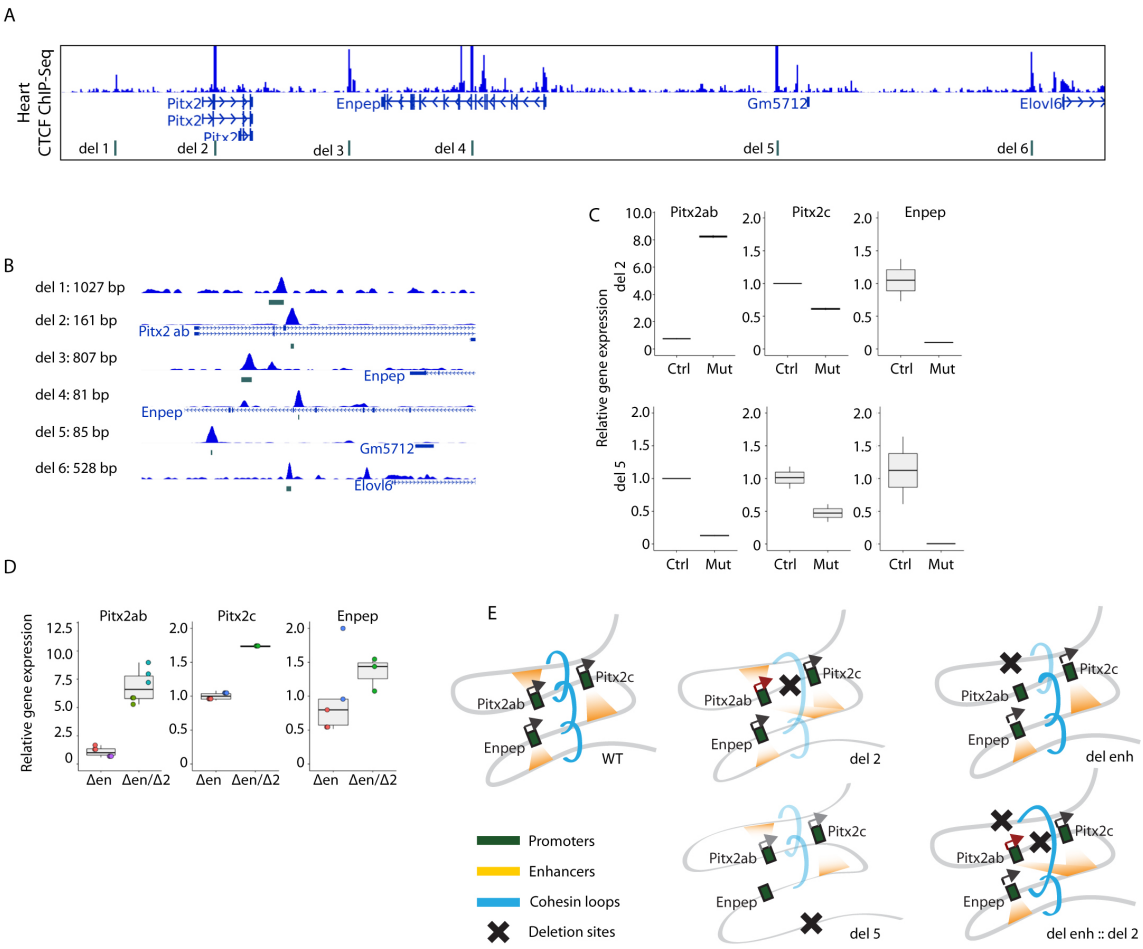


Figure 3.4: CRISPR mediated CTCF sites deletion. (A, B) Targeted CTCF peaks indicated by heart CTCF ChIP-Seq. (C) Gene expression level in mutants of del 2 and del 5. (D) Enhancer / CTCF del 2 double mutants are created using homozygous enhancer mutant line. The deletion of CTCF that segregates Pitx2ab and Pitx2c shows rescue of Pitx2ab level in double mutants. (E) Proposed models show the hypothesis that CTCF controls the specificity of promoter-enhancer interaction. The deletion of CTCF binding sites resets promoter-enhancer looping interaction and destabilizes neighboring looping structures.

3.2 Discussion

Functional analysis of high-risk non-coding variations identified by GWAS is critical for understanding the cause of diseases. In this study, we identified AF-associated region is heart-specific enhancer and directly interact with *PITX2*. Depletion of the enhancer region specifically reduced gene expression of *Pitx2*ab, and potentially destabilized 3D genome structure. The specificity between promoter and enhancer is mediated by genome insulator CTCF protein.

3.2.1 Dynamic of human 4q25 region

Genetic variants on human chromosome 4q25 is highly associated with AF by multiple studies [84]. The most significant SNP, rs2200733, has an odds ratio of 1.9 for the risk allele [57]. Large scale replication study revised 4q25 region and found two additional susceptibility signals marked by rs17579669 and rs3853445 [83]. These SNPs are confined to a 30 kb region and 50 kb away from rs2200733 3.2. Aligned with genomic tracks of DHS and H3K4me1 ChIP-Seq, we identified heart-specific enhancer activity in this region. Curiously, although these SNPs are close to enhancer peaks identified by histone marks, they are not perfectly overlaid. For examples, rs17570669 mismatched with a conserved skeletal-muscle-specific DHS site by 500 bp; rs2200733 is about 5 kb away from closest DHS peak. In some cases, disease-associated variations are located within TF binding motif sequences, which lead to allele-specific TF binding. For example, in ventricle histone 3 lysine 27 acetylation bias towards the G allele against A on rs138143205 locus [70]. In fact, an integrative analysis on haplotype-type-specific epigenome showed 11% (n = 11,714) of all enhancers, are allelic [70] Many of these alleles are disease-associated SNPs identified by GWAS and their function can be tested experimentally in detail. On the other hand, our data indicated a ”multiple enhancer variant” hypothesis, where several variants in linage

disequilibrium impact enhancers and gene expression synergistically [21].

3.2.2 Tissue specific promoter-enhancer pairing

Tissue-specific gene expression pattern is determined by abundance of TFs and their interaction with regulatory elements. *Pitx2ab*, *Pitx2c* and *Enpep*have different expression level across different tissues (Figure 3.5A). For example, in embryonic brain and limb, all three transcripts have relative high expression level; in adult heart, liver and small intestine, *Pitx2ab* has very little expression. To address whether tissue-specific enhancer activity correlates with tissue-specific gene expression in this AF high-risk region. We firstly separated the 200 kb by CTCF binding sites, resulted in two distinct groups: group one including adult small intestine, adult liver and embryonic liver, contains two segregated blocks, marked by CTCF site on *Pitx2* gene body; whereas group two including kidney, heart, limb and brain, contains three blocks, of which the third block was created by CTCF sites in-between *Pitx2* and *Enpep* (Figure 3.5B). We then quantified H3K4me1 intensity in all of the open chromosome, and found a visible correlation between the intensity of the enhancers within the block and the transcription activity of the promoter within the same block. When upstream enhancer were deleted, we found that *Pitx2ab* expression in P19 cells was diminished. However, deletion of CTCF site that segregate *Pitx2ab* and *Pitx2c* promoters restored *Pitx2ab* expression, supporting the hypothesis that CTCF segregates genome into regulatory blocks. Individual block dominates the enhancer-promoter specificity (Figure 3.5C&D).

3.2.3 Potential relevance to AF

We found that AF-associated enhancer specifically regulates *Pitx2ab*, not *Pitx2c in vitro*. Curiously, *Pitx2c* is the dominant isoform that expressed higher level throughout developmental stages in heart [138]. Whether *Pitx2ab* expression is elevated in AF

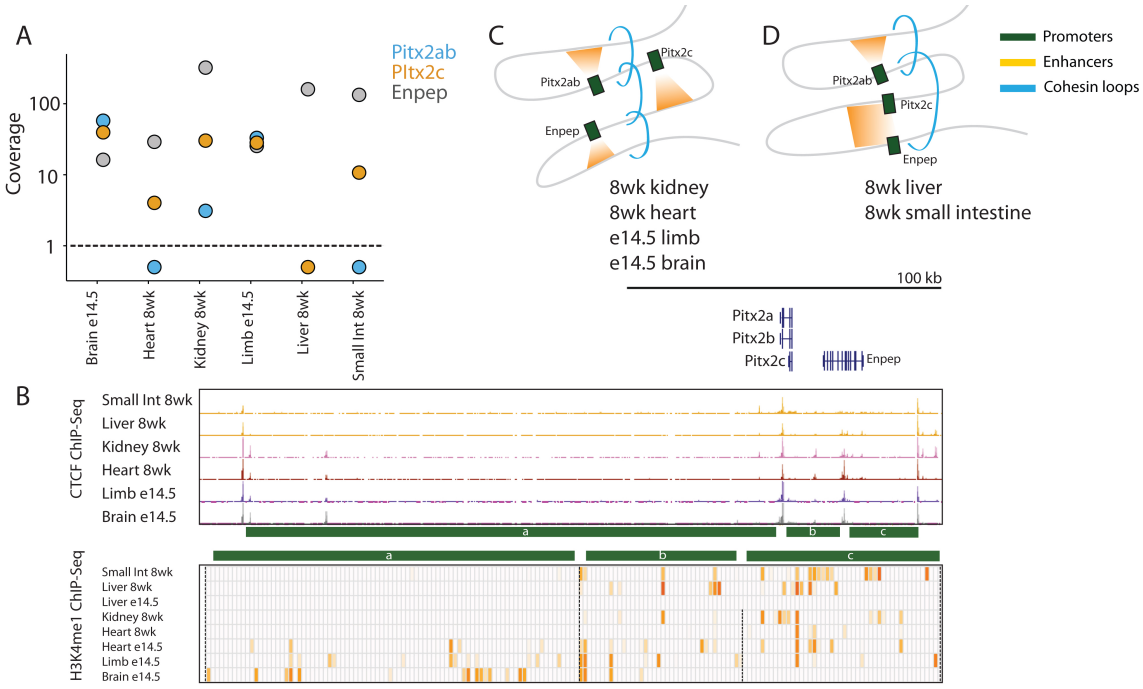


Figure 3.5: Tissue specific promoter-enhancer interactions of *Pitx2ab*, *Pitx2c* and *Enpep*. (A) Gene expression levels of *Pitx2ab*, *Pitx2c* and *Enpep* in different tissues. (B) CTCF and H3K4me1 ChIP-Seq in different tissues. Regulatory regions were segregated by CTCF peaks. For each section, the enhancer activity indicated by H3K4me1 ChIP-Seq intensity is associated with transcription activity. (C) Proposed model for promoter-enhancer specificity in tissues with CTCF peaks between *Pitx2c* and *Enpep* (e.g. kidney, heart, limb and brain); and (D) *Pitx2c* and *Enpep* share the same regulatory section (e.g. liver and small intestine).

Table 3.1: Primers for genotyping of CRISPR-mediated enhancer deletion.

id	Sequences
1f	TTTCTCTTTTGTCACGACTGTAAAGAACAG
1r	AAGTCCATAAGACACACTGAAAACATCTGA
2f	CAAAGATCTTCTTTCCCTATGTTGTGTGTA
2r	TTGGCATAGACTGAGTAGCATCAAGTTAAA

patients is unclear. On the other hand, reactive-oxygen species (ROS) is accumulated in chronicle AF patients [116]. Interestingly, motif analysis of enhancer sequences showed enrichment of consensus binding sites of *MafK* and *Nrf2*, which are known to be activated and shuttled into nucleus upon ROS accumulation [128]. Thus *Pitx2ab* can potentially be activated by *cis*-regulation of *MafK* and *Nrf2* via accumulation of ROS in AF condition.

3.3 Materials and methods

3.3.1 CRISPR-mediated deletion

Two guideRNAs, flanking a 20 kb genomic region, were designed (left: 5'- gtatttcatc-taaggcaca; right: 5'-gatggtcacagtcaatgtg) and were cloned into PX458 vector [107] via golden gate approach. P19 cells were transfected using Lipofectmine 2000 for 24 hrs. Single GFP-positive cells were sorted into 96-well plates and cultured in 10%FBS α-MEM media for seven days. Single clones were then genotyped by using two sets of primers (Table 3.1). Mutant clones were kept for further gene expression analysis.

3.3.2 Chromatin conformation capture sequencing

Detailed protocol of 4C-Seq will be described in 4.3.15. The selected viewpoints are show in (Table 3.2).

Table 3.2: Viewpoints used in 4C-Seq of *Pitx2* promoters and regulatory regions.

chr	Start	End	Description
chr3	128903853	128904369	*Pitx2ab* promoter
chr3	128899347	128899765	*Pitx2ab* promoter
chr3	128916217	128917025	*Pitx2c* promoter
chr3	128915525	128916111	*Pitx2c* promoter
chr3	128867595	128868003	*Pitx2* upstream enhancer
chr3	128832771	128833376	*Pitx2* upstream enhancer
chr3	128834733	128835424	*Pitx2* upstream enhancer
chr3	128873046	128873601	*Pitx2* upstream enhancer
chr3	128959106	128959662	*Pitx2* downstream enhancer
chr3	128222094	128222567	*Pitx2* upstream enhancer
chr3	128222078	128222521	*Pitx2* upstream enhancer

4. ACTIN CYTOSKELETAL REMODELING IS ESSENTIAL FOR HEART REGENERATION IN HIPPO-DEFICIENT MICE*

Although some vertebrates, such as zebrafish, can regenerate the heart, heart regeneration in mammals is limited [61, 11]. Rather than regenerate, human cardiomyocytes undergo a maladaptive stress response commonly termed "pathologic remodeling," including fibrosis and scarring, that leads to heart failure, a leading killer worldwide [132, 90]. The mammalian heart has a transient regenerative capacity that terminates by postnatal day 7 (P7) in mice [102]. This observation has led to the idea that manipulating relevant genetic pathways can therapeutically enhance cardiomyocyte regeneration.

The Hippo signaling pathway is a kinase cascade that links changes in cellular density or mechanical stress to changes in cell proliferation [150]. In mammals, Hippo signaling limits heart size and inhibits cardiomyocyte proliferation during development and adult cardiac regeneration [48, 47]. The downstream Hippo effector Yes-associated protein (Yap) is a transcriptional cofactor that interacts with DNA binding partners such as Tead. When Hippo activity is high, Yap is phosphorylated by Lats and is excluded from the nucleus. When Hippo activity is low, such as during early heart development, Yap shuttles into the nucleus where it promotes cardiomyocyte proliferation[150]. Yap activity is not only regulated by Hippo kinases but also by mechanical signaling. In cells subjected to high amounts of mechanical stress, Yap is preferentially localized in the nucleus and promotes proliferation [89, 6].

Hippo signaling inhibits adult cardiomyocyte regeneration through the Yap effec-

*Parts of this section are from "Actin cytoskeletal remodeling with protrusion formation is essential for heart regeneration in Hippo-deficient mice" by Y. Morikawa, M. Zhang, T. Heallen, J. Leach, G. Tao, Y. Xiao, Y. Bai, W. Li, J. T. Willerson, J. F. Martin. *Science Signaling*, 375(8): ra41, May 2015. Reprinted with permission from AAAS.

tor molecule. Hippo deficiency due to loss of function of the Salvador (*Salv*) adaptor molecule promotes regeneration in mouse models of apex resection, neonatal myocardial infarction, and adult myocardial infarction [47]. Furthermore, overexpression of a constitutively active form of Yap in the myocardium also enhances regeneration in adult cardiomyocytes [148, 75].

In this study, we investigated Yap target genes in heart regeneration by performing ChIP-sequencing (ChIP-Seq) analysis and mRNA expression profiling in Hippo-deficient hearts. Our data indicated that Yap directly regulates genes involved in cell cycle progression, genes that promote F-actin polymerization, and genes linking the actin cytoskeleton to the extracellular matrix.

4.1 Results

4.1.1 Yap directly regulates genes encoding proteins that control cytoskeletal dynamics and cell proliferation

To identify direct transcriptional targets of Yap, we performed ChIP-Seq experiments in *Salv* conditional knockout (CKO) mutant mouse hearts at P8. We chose P8 because it is a non-regenerative stage in wild-type mouse hearts but is a regenerative stage in Hippo-deficient mouse hearts (8). We predicted that in P8 Hippo-deficient hearts, Yap binding would be enriched for genes that are directly involved in cardiac regeneration. We performed ChIP-Seq experiments with an anti-Yap antibody in dissected *Salv* CKO mouse hearts and generated libraries that were sequenced by using an Ion Torrent sequencer [113]. A total of 25 million Yap ChIP-Seq reads were evaluated by using Homer [49]. Motif analysis comparing Yap ChIP-Seq reads indicated that Tead binding elements were among the most enriched peaks, which validated the specificity of the ChIP-Seq experiment (Figure 4.1A&B). In addition, we performed mRNA expression profiling of P8 *Salv* CKO mouse hearts to analyze changes in gene

expression in Hippo-deficient hearts (Figure 4.2). We then compared differentially expressed genes to those in the Yap ChIP-Seq datasets to identify direct transcriptional targets of Yap.

Overlay of the ChIP-Seq and mRNA expression profiling datasets revealed that Yap bound to 928 genes that showed increased expression (Figure 4.1C). From these data, we generated a list of Yap target genes that included 3 gene categories: cell cycle progression, cytoskeleton, and both cell cycle and cytoskeleton (Figure 4.1D-H).

To further characterize productive Yap binding sites, we analyzed our ChIP-Seq dataset for conserved Tead sites and compared our ChIP-Seq data to available DNAase hypersensitivity (DHS) and H3K27Ac datasets that mark enhancers [96, 93, 12]. Comparison of a mouse heart H3K27Ac ChIP-Seq dataset that marks enhancers with the Yap ChIP-Seq dataset showed that many Yap peaks from Hippo-deficient hearts were enriched in putative enhancer regions [96] (Figure 4.3A-E).

We validated the transcriptional activity of a subset of genes by using luciferase assays. We evaluated genes encoding proteins that are involved in both cytoskeletal remodeling and the linking of the cytoskeleton to the extracellular matrix (ECM), such as *Actrt2*, *Pkp4*, *Enah*, *Fmn2*, *Sgcd*, and *Ctnna3*. We also tested genomic fragments from the cell cycle genes *Aurkb* and *Lin9*. A luciferase reporter gene was fused to amplified genomic regions containing conserved Yap ChIP-Seq peaks that aligned with DHS peaks and/or H3K27Ac peaks that also contained Tead binding elements [96, 93] (Figure 4.3A-D). We also performed chromosome conformation capture (4C) analysis to define enhancer contact maps for *Sgcd* and *Ctnna3* and compared those maps to Yap ChIP-Seq data (Figure 4.3C&D). Chromosome conformation capture provides 3-dimensional information about enhancer-promoter interactions and can aid the discovery of previously unknown enhancers [37, 23]. We noted that enhancer-promoter contact points for both genes included Yap ChIP-Seq peaks, further supporting the con-

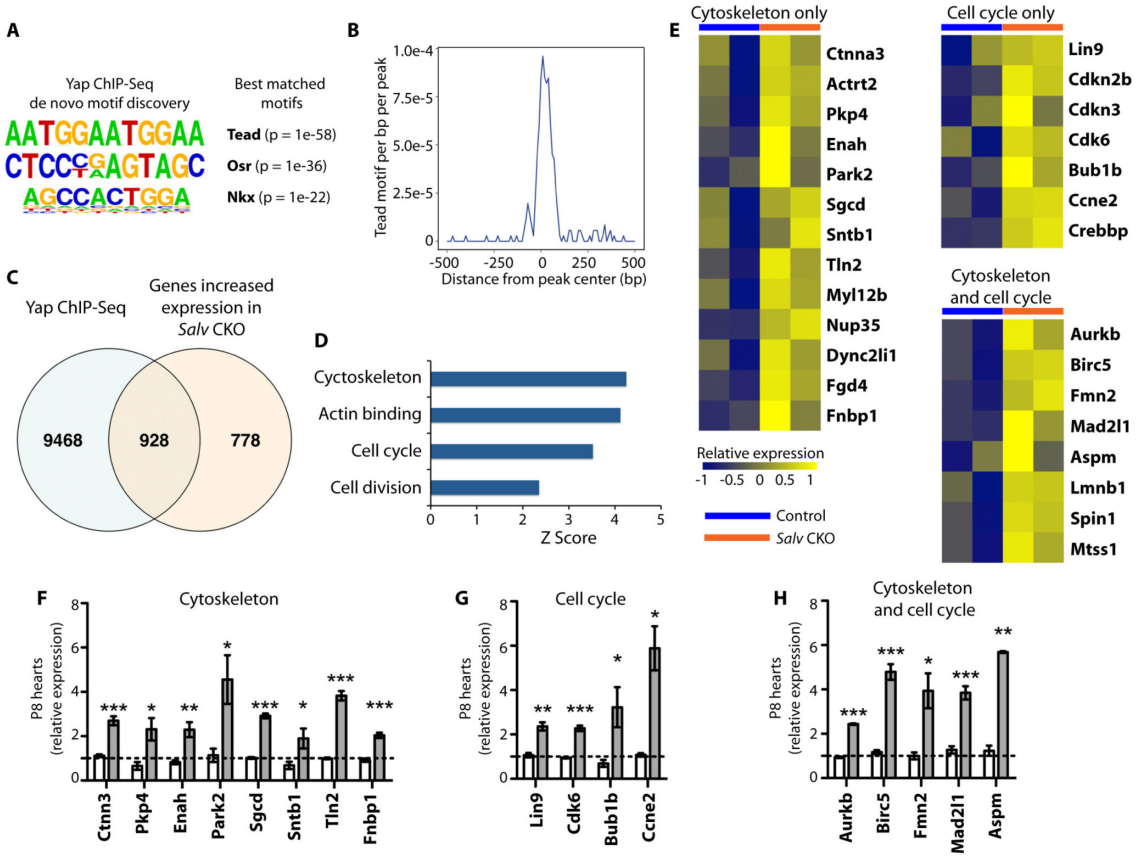


Figure 4.1: Integrated genomic analysis for identifying Yap target genes. (A) Motif analysis for enriched Yap ChIP-Seq peaks (total number = 35,412 from 2 independent biological replicates). *De novo* motifs and their best matches are shown. (B) The density of Tead binding motifs within a 500-bp distance of the Yap ChIP-Seq peaks is shown. (C) Overlay of genes with increased expression in *Salv* CKO mouse hearts (total number = 1,706 from 2 mice) and genes annotated from Yap ChIP-Seq peaks (total number = 10,396). (D) Gene ontology analysis of genes with increased expression in Salv CKO mouse hearts and with Yap binding peaks (total number = 928). Enriched terms were calculated by using over-representation statistics and measured by using *Z*-scores. (E) Heat map of Yap target genes identified by the overlay of microarray and Yap ChIP-Seq genes in the labeled categories. Heat maps show relative expression between *Salv* CKO and control mouse hearts. (F-H) qRTPCR validation of Yap target genes in P8 control (unshaded bar) and P8 Salv CKO (shaded bar) mouse hearts. $n = 3$ independent biological replicates. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

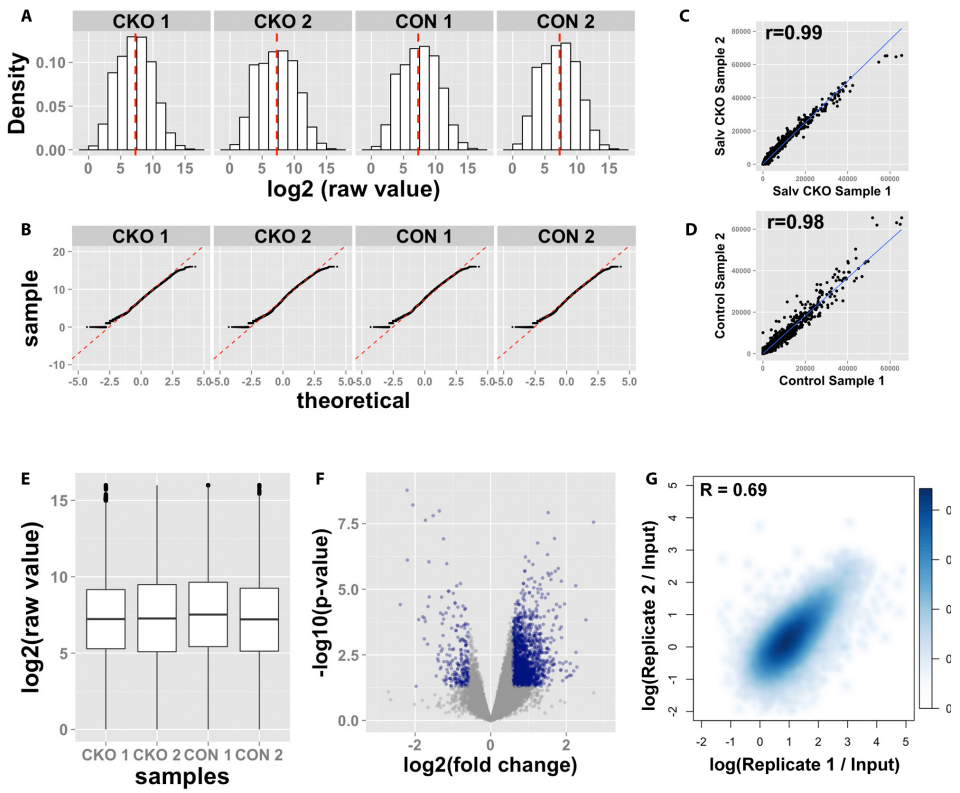


Figure 4.2: Microarray and Yap ChIP-Seq reproducibility. Microarray was performed on P8 control and *Salv* CKO mouse hearts (2 biological replicates per group). Each sample has 3 technical controls. (A) Histograms show the distribution of log-transformed raw signal values which are approximate to a normal distribution. The red dotted lines indicate the mean value. (B) Quantile-Quantile plot of log2 (raw signal value) from control and *Salv* CKO samples. The *x*-axes show quantile values of standard normal distribution. The *y*-axes show quantile value of log-transformed signal value. (C, D) Reproducibility of microarray raw signal value in control and *Salv* CKO samples. Biological replicates were plotted as x-axis and y-axis values individually. Linear correlation coefficients were measured, as *r* is given in each diagram. The blue lines indicate the line of best fit. (E) Boxplots show that there were no significant outliers among the samples. (F) Differential expressed genes in control and *Salv* CKO samples. Individual genes were plotted by using log of fold-change as the *x*-axis and -log10 of the *p*-value as the *y*-axis index. Differential expression threshold was fold change of at least 1.5 and *p*-value of at least 0.05, highlighted in blue. (G) Reproducibility of Yap ChIP-Seq (2 biological replicates). Signal intensities were measured over input sample in 5 kb bins. Pearson's correlation coefficient was calculated based on the basis of log-transformed values.

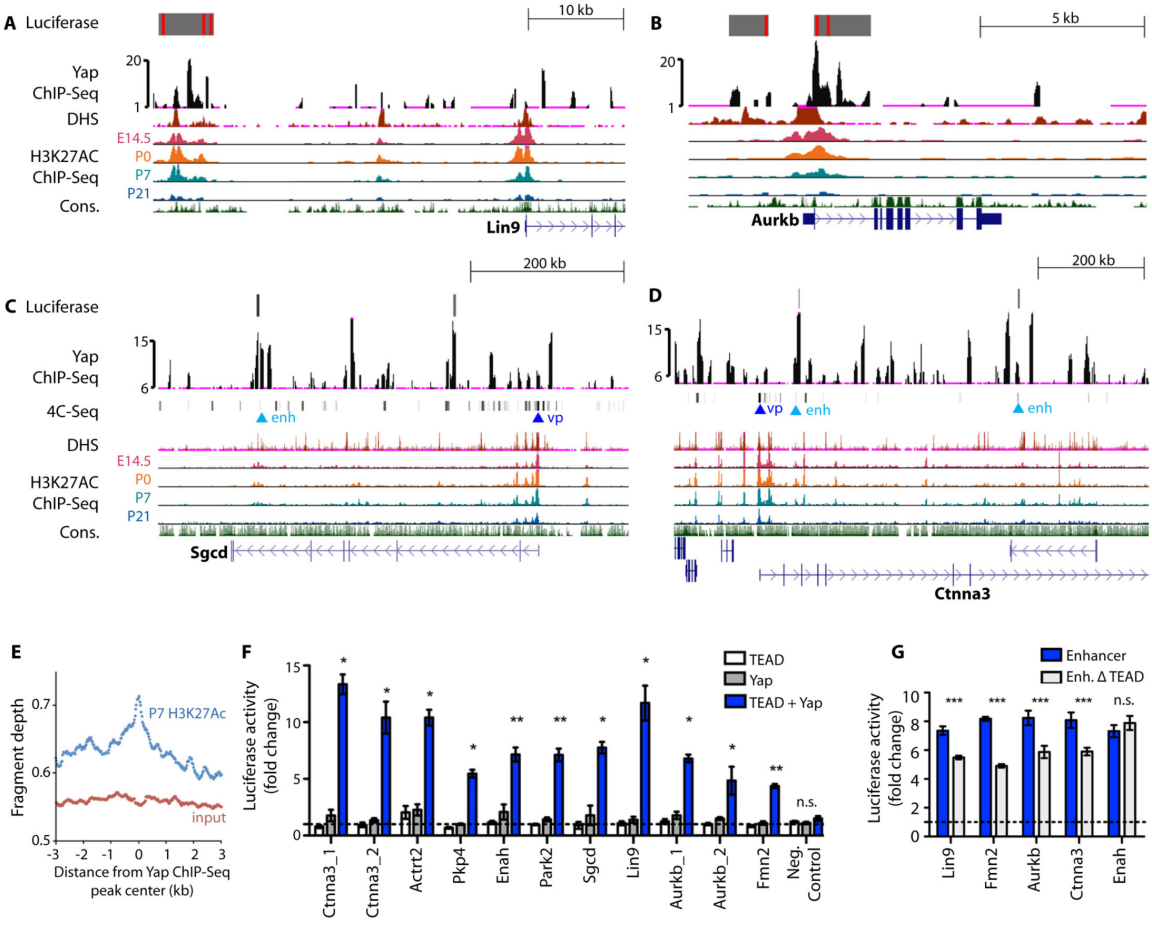


Figure 4.3: Preferential expression of Yap target genes in the fetal heart. (A-D) Genome browser view of Yap ChIP-Seq enriched peaks for the labeled genes. Alignments are shown for stage-specific H3K27Ac ChIP-Seq, DHS, and 4C anchor points (*Ctnna3*, *Sgcd*). For 4C, vp denotes viewpoint and enh denotes enhancer. Grey blocks show regulatory regions used as luciferase reporters with red lines indicating consensus Tead binding motif. The *y*-axes of Yap ChIP peaks shows the normalized read number. (E) Quantification of P7 heart H3K27Ac ChIP-Seq reads in the 6-kb range around Yap ChIP-Seq peaks. $n = 2$ biologic replicates (F) Luciferase reporter assay data showing that Yap/Tead co-activated target gene expression through regulatory regions identified in Yap ChIP-Seq. $n = 3$ independent transfection experiments. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. (G) Luciferase reporter assay data of Yap enhancers with or without TEAD motifs. $n = 3$ independent transfection experiments. All luciferase constructs were co-transfected with Yap and TEAD expression vectors. *** $p < 0.001$ (Mann-Whitney), *n.s.* not significant. Error bars are standard deviations. Activity was normalized to that of pGL3 vector.

clusion that Yap peaks were found in enhancer regions. Thus, we generated luciferase reporter constructs containing the Yap peaks that were also 4C contact points.

The transfection of cells with a Yap expression vector trans-activated enhancer elements of these genes in vitro in a manner that required Tead2 co-transfection (Figure 4.3F&G). Moreover, deletion of the Tead elements resulted in loss of enhancer activity in all genes with the exception of *Enah*. It is possible that *Enah* has a poorly conserved Tead element that we failed to identify or that Yap may bind *Enah* by cooperating with a different DNA binding partner (Figure 4.3F&G). Together, these data support the hypothesis that Yap directs the transcription of genes promoting cell cycle progression, genes that link the actin cytoskeleton to the ECM, and genes that remodel the actin cytoskeleton.

4.1.2 Yap target genes are preferentially expressed in the fetal heart

We analyzed human RNA-Seq data to determine whether orthologs of our Yap target genes are more highly expressed in human fetal or adult hearts. This analysis indicated that the expression of most Yap target gene orthologs was higher in human fetal hearts than in adult hearts [12] (Figure 4.4A&B). Comparison of the human fetal RNA-Seq data with genes that we found had increased expression in *Salv*CKO hearts indicated an enrichment in fetal hearts for genes that encode cell cycle and cytoskeletal proteins (Figure 4.4C). In addition, genes encoding proteins typical of the adult heart, including lipid metabolism and muscle contraction proteins, had reduced expression in Salv CKO mouse hearts (Figure 4.4D). This is consistent with findings that developing and early postnatal hearts rely on glycolytic metabolism rather than lipid oxidation and have an immature contractile apparatus [82]. We also compared our Yap target gene set to mRNA expression data from P4 rat cardiomyocytes overexpressing Yap. Almost all Yap targets we identified had higher mRNA levels in the Yap gain of function rat

cardiomyocytes (fig. S2E) [134].

Analysis of DHS datasets from human fetal and adult hearts further supported the hypothesis that Yap binding sites are enriched in fetal heart genes [88, 130]. Motif analysis from the human fetal and human adult DHS data indicated that Tead elements were highly enriched in human fetal heart DHS peaks, along with recognition elements for other important cardiac development transcription factors such as Tbx and Mef2 (Figure 4.5A). Although Tead elements were enriched in human fetal heart DHS peaks, this was not the case for DHS peaks from human embryonic stem cells, supporting the cardiac tissue enrichment of Tead binding activity (Figure 4.5B). Tead elements were also found in DHS peaks from tissues such as the kidney and lung indicating that Tead elements are not cardiac specific (Figure 4.5C). Tead elements were enriched in fetal heart DHS peaks even for genes whose physical location was not conserved between humans and mice (Figure 4.5D-F). These data support the hypothesis that Yap and Tead enhance the expression of fetal cardiac genes.

4.1.3 Hippo-deficient adult cardiomyocytes are proliferative

We evaluated proliferation and apoptosis in adult and neonatal control and *Salv* CKO mutant mouse hearts after inducing myocardial infarction by left anterior descending artery occlusion (LADO; Figure 4.6). Compared to controls, *Salv* CKO mutant mouse hearts showed an increased number of EdU-positive cardiomyocytes at 4, 10, and 15 days post myocardial infarction for adults (Figure 4.6A-C) and 1 day post myocardial infarction for P8 neonatal hearts (Figure 4.6C). Notably, S-phase entry was induced by Hippo deficiency and not by injury, as shown by equivalent amounts of EdU incorporation in Hippo-deficient sham hearts and Hippo-deficient injured hearts (Figure 4.6C). The number of M-phase cardiomyocytes, as indicated by Aurora B immunostaining, was higher in Hippo-deficient injured hearts than in Hippo-deficient

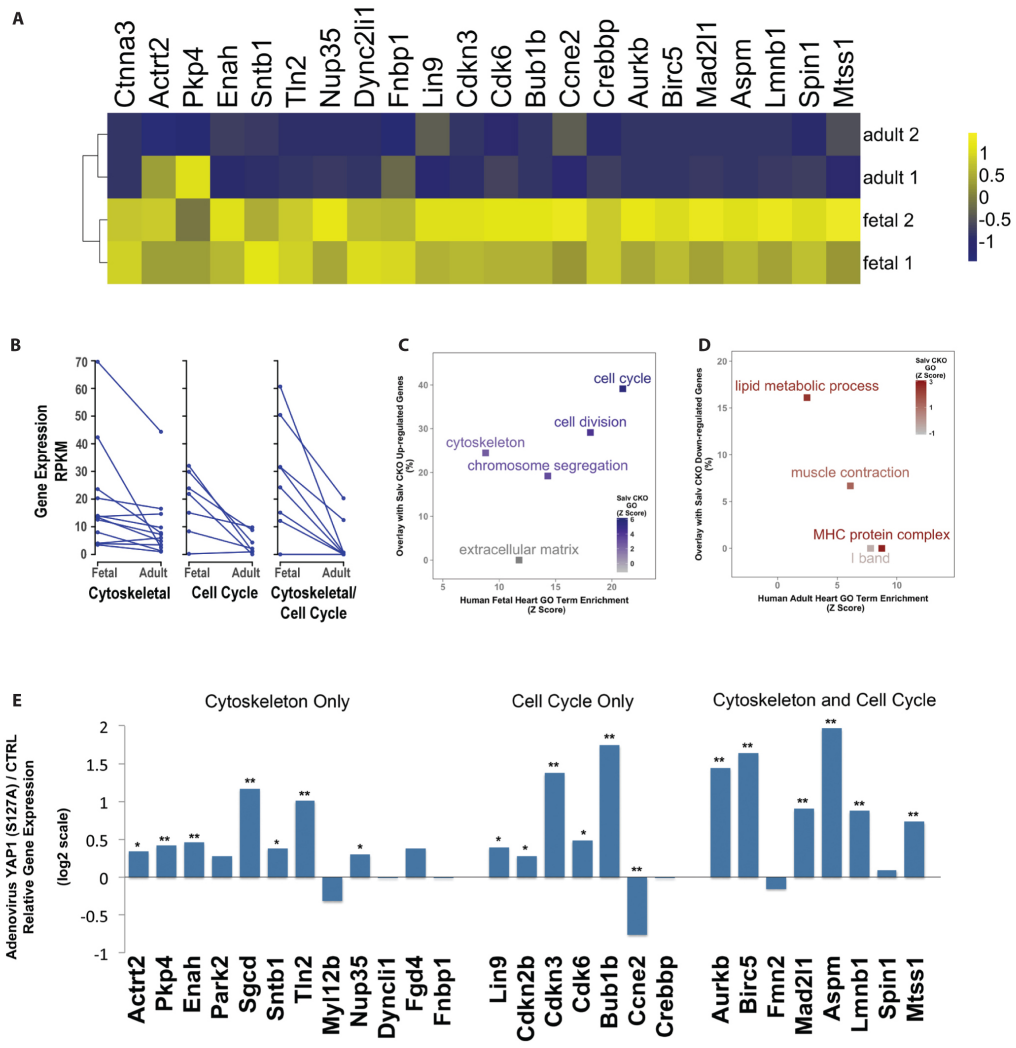


Figure 4.4: Yap target genes are enriched in the fetal heart. (A) Heat map of Yap target gene expression in human fetal and adult ventricle (n = 2 biological replicates for each). The color bar represents relative expression level of log-transformed value for each gene across different samples. (B) Yap target gene orthologs are more highly expressed in human fetal hearts than in adult hearts. Gene expression was measured in RPKM by using RNA-Seq profiling data. n = 2 independent biological replicates. (C, D) Gene ontology analysis for genes enriched in fetal (C) (total gene number = 700) and adult hearts (D) (total gene number = 417). x-axes indicate Z-score of over-representation statistics; y-axes indicate percentage of gene overlay with differentially expressed genes from Salv CKO hearts. (E) Target gene expression in postnatal day 4 neonatal rat cardiomyocytes transduced with lacZ or YAP1 (S127A) expressing adenovirus (data extracted from GSE33019, n = 4 biological replicates per group. ** false discover rate (fdr) < 0.01; * fdr < 0.05.

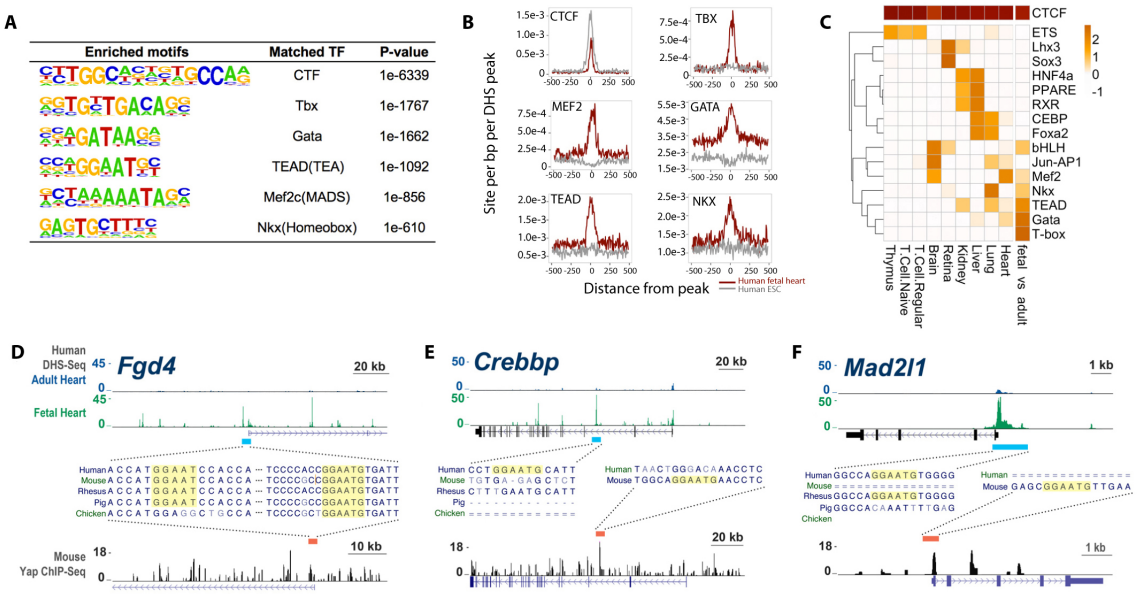


Figure 4.5: Yap/Tead regulatory elements are enriched in fetal heart DNaseI hypersensitivity sites (DHS) and Yap binding peaks. (A) Matched motifs in human fetal heart DHS-Seq peaks (n = 11 biological replicates). Human fetal heart-specific DHS peaks were compared to human adult heart DHS peaks (n = 2 biological replicates). Enriched motifs represent fetal heart-specific DHS peaks. (B) The density of enriched motifs within a 500 bp range of human fetal heart DHS peaks are shown in comparison with that of human embryonic stem cell DHS peaks (grey). (C) Heat map of enriched motifs in DHS footprints from different adult tissues and fetal vs. adult comparison. The relative enrichment was measured by the number of footprints containing each motif. (D-F) Representative genome browser tracks of Yap targets in regions of genome that are not conserved between species. DHS-Seq peaks from human fetal (green) and adult (blue) hearts and Yap ChIP-Seq peaks from mouse P8 *Salv* CKO hearts (black) are shown. *Fgd4* contains highly conserved Tead binding elements in the promoter region (D). In *Crebbp* (E) and *Mad2l1* (F), fetal specific DHS peaks and Yap binding peaks are located in genomic regions that are not conserved between species (non-syntenic).

sham hearts (Figure 4.6D-G). This indicates that Hippo-deficient hearts respond to injury by progressing through cytokinesis more efficiently than do Hippo-deficient sham hearts.

The number of apoptotic cells in the myocardium was similar between control and *Salv* CKO mice after LADO (Figure 4.6H). Moreover, there was no evidence for a higher frequency of cell fusion events in *Salv* CKO mutant mouse hearts.We also examined cell size at multiple stages after myocardial infarction and found that *Salv* CKO mouse cardiomyocytes were smaller than those of controls. We conclude that *Salv* CKO mouse cardiomyocytes showed increased EdU incorporation for 15 days post myocardial infarction and that after injury, *Salv* CKO mouse cardiomyocytes progress through M-phase more efficiently than do cardiomyocytes of sham mouse hearts.

In the border zone of *Salv* CKO mouse hearts, nuclear localized Yap was increased (Figure 4.6I-M), suggesting the possibility that the expression of Yap-regulated genes may also be increased in the border zone of the regenerating heart. To examine this possibility, we isolated RNA from dissected border zones of *Salv* CKO mouse hearts and control hearts 4 days after inducing myocardial infarction. Quantitative real-time PCR (qRT-PCR) analysis revealed that Yap target genes had increased expression in the border zone of *Salv*CKO mouse hearts as compared to that of controls (Figure 4.6N).

4.1.4 Hippo-deficient cardiomyocytes extend sarcomere-filled protrusions

Our finding that Yap-regulated genes encode proteins that regulate the actin cytoskeleton suggests that Hippo-deficient cardiomyocytes may have different cytoskeletal characteristics than control cardiomyocytes. We examined cardiomyocytes at 4, 7, and 10 days after myocardial infarction in adult control and *Salv* CKO mouse hearts. No differences in cellular morphology were observed between control and *Salv* CKO

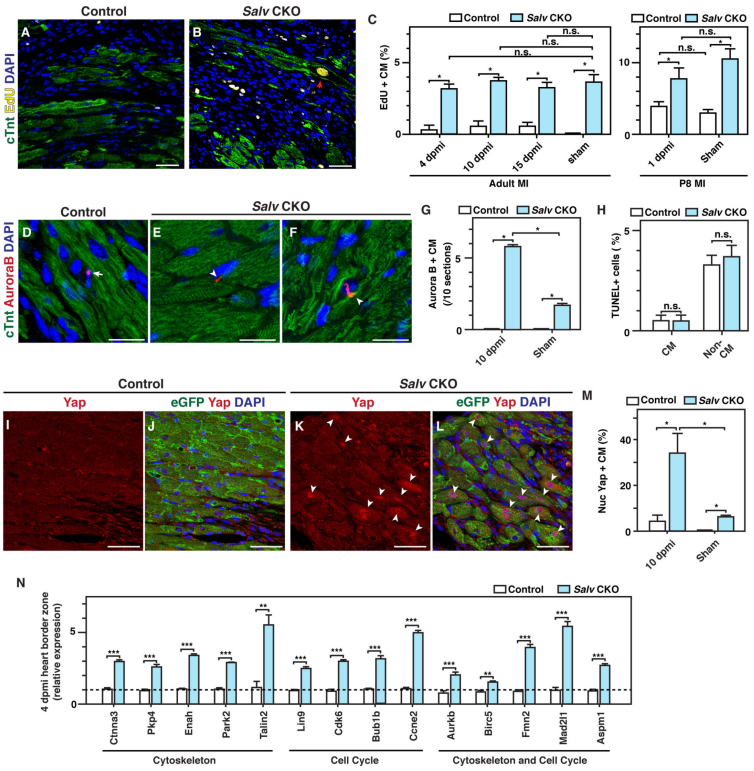


Figure 4.6: DNA synthesis and Yap localization in border zone cardiomyocytes during adult heart regeneration. LAD occlusion was performed on *Myh6-CreErt; mTmG* (control) and *Myh6-CreErt; Salvfx/fx; mTmG* (*Salv*CKO) mice, and hearts were collected at 1, 4, 10, and 15 days post myocardial infarction (dpmi). (A, B) *De novo* DNA synthesis was detected by measuring EdU incorporation in control (A) and *Salv*CKO (B) mouse hearts at 10 dpmi. Red arrowhead shows EdU-stained nucleus. Bars = 50 µm. (C) Quantification of *de novo* DNA synthesis in the border zone at 1, 4, 10, and 15 dpmi or in sham mice (n = 3 independent biological replicates for each genotype and time point). *$p < 0.05$. (D-F) Cytokinesis was detected with Aurora B kinase immunostaining in control (D) and Salv CKO (E, F) mouse hearts at 10 dpmi. Arrow shows staining in non-cardiomyocyte, and arrowheads show staining in cardiomyocytes. Bars = 20 µm. (G, H): Quantification of Aurora B kinase (G) and TUNEL activity (H) (n = 3 independent biological replicates for each genotype). ** p <0.01. (I-M) Yap localization in border zone of control (I, J) and Salv CKO mouse hearts (K, L) at 10 dpmi. Arrowheads show nuclear localized Yap. Bars = 50 µm. (M) Quantification of nuclear Yap in border zone cardiomyocytes at 10 dpmi (n = 3 independent biological replicates for each genotype). * $p < 0.05$; remaining column comparisons were nonsignificant. (N) Gene expression of Yap downstream target genes were quantified with qPCR in border zones from heart tissues after myocardial infarction. Border zone is defined as tissue adjacent to the scar (n = 3 biologic replicates) ** p <0.01, ***p <0.001 Figure credit: Yuka Morikawa, Todd Heallen

mouse hearts until 7 days post myocardial infarction, when the *Salv* CKO mouse cardiomyocytes showed more prevalent sarcomere breakdown in the border zone (Figure 4.7A-G). At 10 days post myocardial infarction, adult *Salv* CKO mouse cardiomyocytes, but not those from control mice, showed extended protrusions into the scar (Figure 4.7C, F, &G). Immunofluoresence experiments performed by using anti-talin and anti-vinculin antibodies that recognize costameres linking the ECM to the actin cytoskeleton through the integrin-vinculin-talin complex revealed that the distribution of both talin and vinculin [153] was markedly expanded in the cardiomyocytes of adult *Salv* CKO mice at 10 days post myocardial infarction (Figure 4.7H-O). Protrusions were commonly seen on multiple sections surrounding the infarct in adult *Salv* CKO mouse cardiomyocytes but not in those of control mice.

In the apex resection model, a defined region of cardiac apex is removed, allowing for the accurate characterization of renewing cardiomyocytes [102]. Cells derived from the cardiomyocyte lineage were observed in the resected zone that contained a large number of non-cardiomyocyte cells. Similar to what we observed in adult mouse hearts after myocardial infarction, Hippo-deficient cardiomyocyte lineage–derived cells extended cellular protrusions and showed evidence for extensive remodeling of vinculin-positive focal adhesions. In control mouse hearts, border-zone GFP-positive cells did not infiltrate the resected region of the heart or remodel vinculin.

4.1.5 Hippo-deficient cardiomyocytes mobilize into collagen gels

Our ChIP-Seq and immunofluorescence data support the notion that *Salv* CKO mouse cardiomyocytes display extensive cytoskeletal remodeling with protrusive activity. Because cellular protrusions are characteristic of migrating cells [109], we determined whether *Salv* CKO mouse cardiomyocytes are capable of mobilizing into a collagen gel. Cardiomyocytes from cardiac explants from *Salv* CKO mice, but

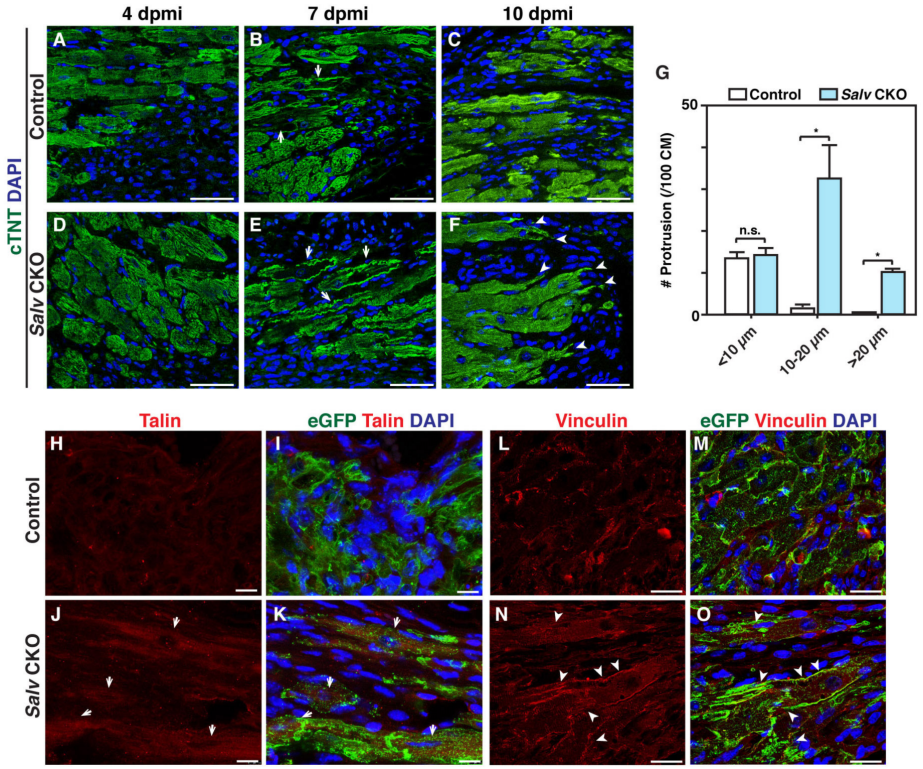


Figure 4.7: Cardiomyocyte morphological change and cytoskeleton rearrangement during adult heart regeneration. LAD occlusion was performed on the hearts of *Myh6-CreErt; mTmG* (control) and *Myh6-CreErt; Salvfx/fx; mTmG* (*Salv*CKO) mice. Control (A-C) and *Salv*CKO (D-F) mouse hearts were stained for the cardiomyocyte marker cTNT to visualize morphology of the cardiomyocytes in the border zone at 4, 7, and 10 dpmi (n = 3 independent biological replicates for each genotype and time point). Arrows show cardiomyocytes with sarcomere disassembly. Arrowheads show cardiomyocyte protrusion. Bars = 50 μm. (G) Quantification of protrusions. Cardiomyocytes adjacent to the scar were analyzed for length and number of protrusions at 10 dpmi. One hundred cardiomyocytes from each biological replicate were analyzed (n =3 independent biological replicates). * $p < 0.05$. Control (H, I) and *Salv*CKO (J, K) mouse heart sections at 10 dpmi were stained for *Talin* (n = 2 independent biological replicates). Arrows show increased Talin staining in border zone cardiomyocytes. Bars = 25 μm. Control (L, M) and *Salv*CKO (N, O) mouse heart sections at 10 dpmi were stained for the focal adhesion molecule vinculin. Arrowheads show the rearrangement of vinculin in the protruding front of the cardiomyocytes (n = 3 independent biological replicates). Bars = 25 μm. Figure credit: Yuka Morikawa, John Leach

not those from control mice, efficiently migrated into collagen gels (Figure 4.8A-E). Therefore, our findings suggest that *Salv*CKO mouse cardiomyocytes extend protrusions that are sufficient for cellular movement into a collagen gel. We also used the collage invasion assay to test whether the small interfering (si) RNA-mediated knockdown of Salv promotes cell migration into the gel. siRNA knockdown experiments in P19 embryo carcinoma cells revealed that the siRNA-mediated knockdown of Salv resulted in increased cell invasion into the gel that was not due to increased cell proliferation (Figure 4.8F-L). In addition, the siRNA-mediated knockdown of the Yap target genes *Fgd4*, *Pkp4*, and *Talin2* resulted in decreased cell migration into collagen gels. Furthermore, the simultaneous siRNA-mediated knockdown of both Salv and Talin2 showed that the knockdown of Talin2 suppressed the Salv knockdown phenotype. Moreover, treating P19 cells with verteporfin, a small molecule that disrupts Yap-Tead interaction, blocked the migration of P19 cells into the gel (Figure 4.8L) [81].

4.1.6 The dystrophin glycoprotein complex is required for cardiac regeneration

Many Yap target genes encode proteins that link the actin cytoskeleton to the sarcolemma and ECM. Sarcoglycan delta (*Sgcd*) and Syntrophin B1 (*Sntb1*) are both DGC components that connect the actin cytoskeleton to the extracellular matrix and transmit force from muscle cells to the extracellular matrix [9, 39]. *Sgcd* is mutated in human patients with limb-girdle muscular dystrophy and is thought to stabilize the plasma membrane in response to mechanical stress [135]. qRT-PCR and Western blot analyses showed that the abundance of Sgcd mRNA and Sgcd protein was increased in *Salv*CKO mouse hearts when compared with control mouse hearts (Figure 4.9A&B). After myocardial infarction, Sgcd mRNA levels were increased in both control and *Salv*CKO mouse hearts. The abundance of Sgcd protein was not further increased af-

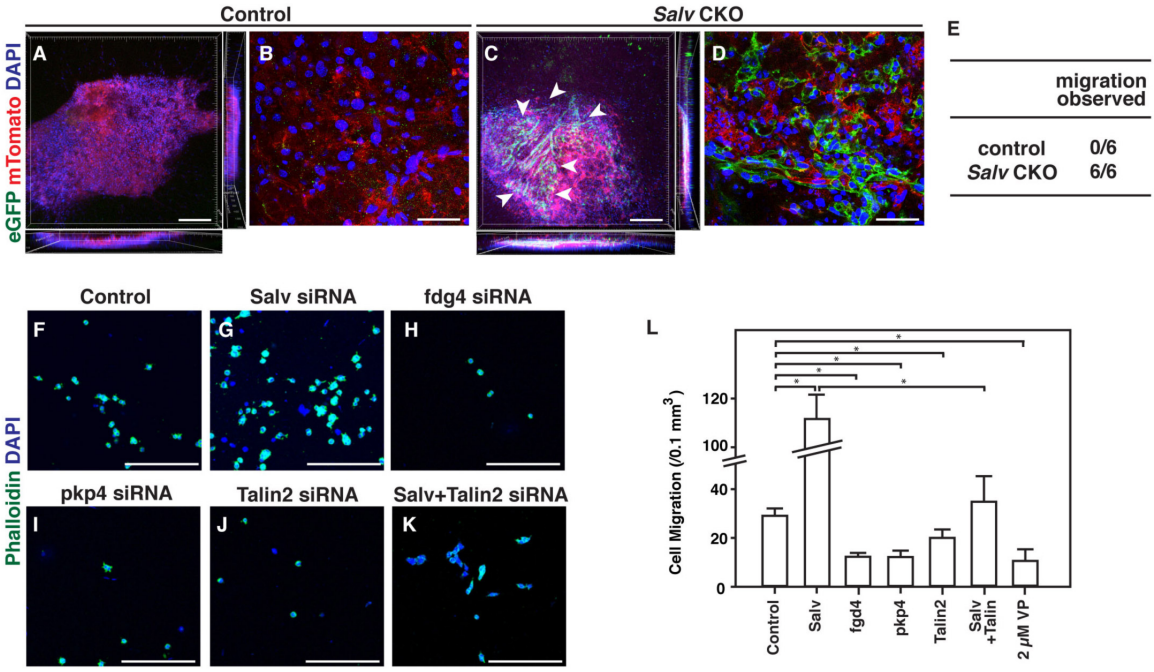


Figure 4.8: Cardiomyocyte migration through collagen. Collagen migration assays were performed with P8 cardiomyocytes from control (*Myh6-CreErt; mTmG*)(A, C) or *Salv*CKO (*Myh6-CreErt; Salvfx/fx; mTmG*) mouse hearts (B, D). Cardiomyocyte lineage was visualized with eGFP and non-cardiomyocytes were visualized with mTomato. (A, C) Whole gel view. Arrowheads show migrated cardiomyocytes. Bars=250 μm. (B, D) High-magnification images of the area with eGFP positive cells. Bars=50 μm. (E) Quantification of the number of hearts in which migration was observed for each genotype ($n = 6$ hearts per genotype). $p = 0.002$, control *vs.* *Salv*CKO mice. (F-K) P19 cell migration in collagen gel after siRNA treatment.with the labeled siRNA Bars = 250 μm. (L) Quantification of migrated cells after each treatment. Cells were treated with either siRNA or Yap inhibitor verteporfin (VP). $n = 3$ biologic replicates for all groups. Figure credit: Todd Heallen, Yuka Morikawa

ter myocardial infarction, suggesting the posttranscriptional regulation of *Sgcd* (Figure 4.9A&B). Western blot analysis indicated that Sntb1 protein abundance was unchanged between control and *Salv*CKO mouse hearts. We next fused the region of the *Sgcd* gene containing the Yap/Tead binding element to a luciferase reporter gene. Luciferase assays revealed that Yap and Tead2 transactivated the *Sgcd* reporter gene but not a Sgcd reporter with a mutated Tead element (Figure 4.9C).

Yap is required for mammalian heart regeneration during the postnatal regenerative period [148]. Therefore, we examined whether the Yap target gene *Sgcd* and its functional activity in the DGC are also required for cardiac regeneration. As a model for impaired DGC, we used *Mdx* mutant mice that harbor a null dystrophin mutation that disrupts the DGC. Cardiac apexes from *Mdx* and control mice were amputated at P1, when regeneration is possible [102]. Whereas control mouse hearts had a robust regenerative capacity, all *Mdx* mouse hearts formed a large scar (Figure 4.9D-F). Furthermore, cardiac function as measured by ejection fraction and fractional shortening was compromised in resected *Mdx*mouse hearts (Figure 4.9G).

4.1.7 The dystrophin glycoprotein complex is required for cellular protrusions

We examined the border zone of control and *Mdx*mouse hearts 4 days after resection. No difference was observed in the amount of nuclear localized Yap between *Mdx*and control mouse hearts, suggesting that DGC function is dispensable for the shuttling of Yap from the nucleus to the cytoplasm during regenerative stages (Figure 4.10A-E). Whereas cardiomyocytes of control mice exhibited extensive cytoskeletal remodeling with protrusion formation, *Mdx*mouse cardiomyocytes failed to remodel their cytoskeleton and extend protrusions (Figure 4.10F-H). Using collagen gel assays to functionally evaluate protrusive activity, we found that explanted myocardial cells from control mice were competent to move through a filter into a lower collagen

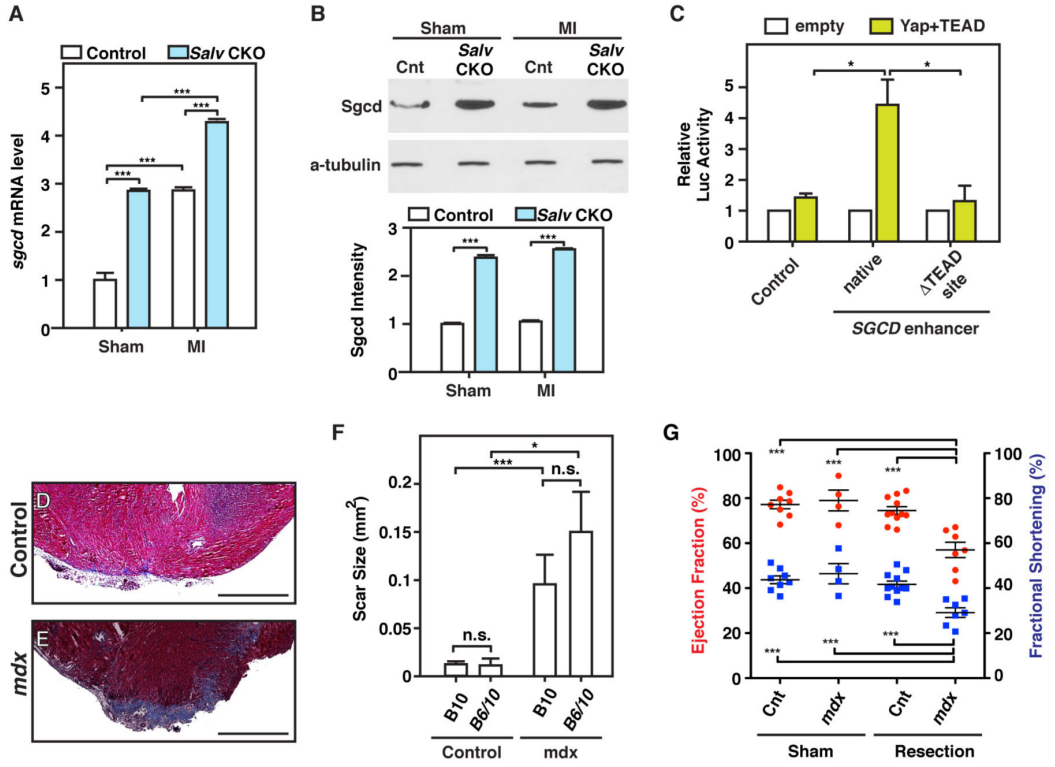


Figure 4.9: The dystrophic complex is downstream of the. Hippo pathway and is required for cardiac regeneration (A, B) LADO and sham surgery were performed in control and *Salv*CKO mouse hearts at P8, and heart samples were collected at 4 dpmi. (A) Delta-sarcoglycan (*Sgcd*) mRNA was detected in control and *Salv*CKO mouse hearts by using qRT-PCR and was quantified by normalization to *Gapdh* (n = 3 for all groups). (B) Sgcd protein was detected by using Western blot analysis (n = 3 hearts per genotype and treatment). Sgcd band intensities were quantified and normalized to those of alpha-tubulin. ***$p < 0.001$, remaining column comparisons were non-significant (*n.s.*) (C) Luciferase assays were performed with P19 embryonal carcinoma cells. Cells were transfected with either the control luciferase reporter, a reporter containing the Sgcd enhancer, or a reporter containing the *Sgcd* enhancer but lacking the Tead site. Three independent experiments with technical triplicates were performed. *$p < 0.05$. (D-G) Dystophin glycoprotein complex (DGC) is required for endogenous cardiac regeneration. (D, E) Representative images of trichrome-stained heart sections from B10 control (D) and *Mdx*-B10 (E) mice subjected to resection of the cardiac apex. Bars = 500 µm. (F) Quantification of the scar size at 21 dpr in B10, (n = 11), B6/10 (n = 4), *Mdx*-B10 (n = 7), and *Mdx*-B6/10 (ittn = 6) mouse hearts.*$p < 0.05$, ***$p < 0.001$. (G) Echocardiography analysis of control sham (n = 3), control apex resection (n = 7), *Mdx*sham (n = 4), and *Mdx*apex resection (n = 7) mouse hearts 21 days after surgery. *** $p < 0.001$, remaining column comparisons were *n.s.* Figure credit: John Leach, Todd Heallen

gel, whereas myocardial cells from *Mdx*mice failed to do so (Figure 4.10I-L). We also performed EdU incorporation and AurkB immunofluorescence assays to determine whether *Mdx*mouse cardiomyocytes had a proliferative defect that was responsible for failed regeneration. We found that EdU incorporation and AurkB immunoreactivity were equivalent between cardiomyocytes of *Mdx*and control mice. Proliferation was increased equally in both control and *Mdx*mouse hearts after injury, supporting the notion that *Mdx*mice failed to regenerate the heart because of a defect in myocardial protrusive activity rather than because of a proliferative defect. To determine whether dystrophin inhibition could suppress the increased cell migration phenotype observed in Salv knockdown cells, we concomitantly knocked down *Salv*and *Dmd* in p19 cells and evaluated cell migration. Indeed, knocking down both *Dmd* and *Salv*in p19 cells suppressed the increased cell migration observed in *Salv*knockdown cells (Figure 4.10M-Q).

4.2 Discussion

Mobilization of endogenous cardiomyocyte regenerative capacity may be a viable option for definitive heart failure therapy. Hippo pathway loss of function or Yap gain of function promotes cardiomyocyte regeneration after myocardial infarction [47, 148, 75]. In this study, we identified direct Yap transcriptional target genes and discovered the importance of cytoskeletal remodeling with cellular protrusions in heart regeneration. Yap target genes directly control cell shape and protrusive activity.

4.2.1 Yap regulates genes encoding proteins that link the ECM to the actin cytoskeleton

Cardiomyocytes constantly sense and interpret local microenvironment rigidity [66]. Yap-regulated genes, including *Ctnna3*, *Sgcd*, and *Tln2* are important for linking the actin cytoskeleton to the ECM. Our data, along with data showing that Yap

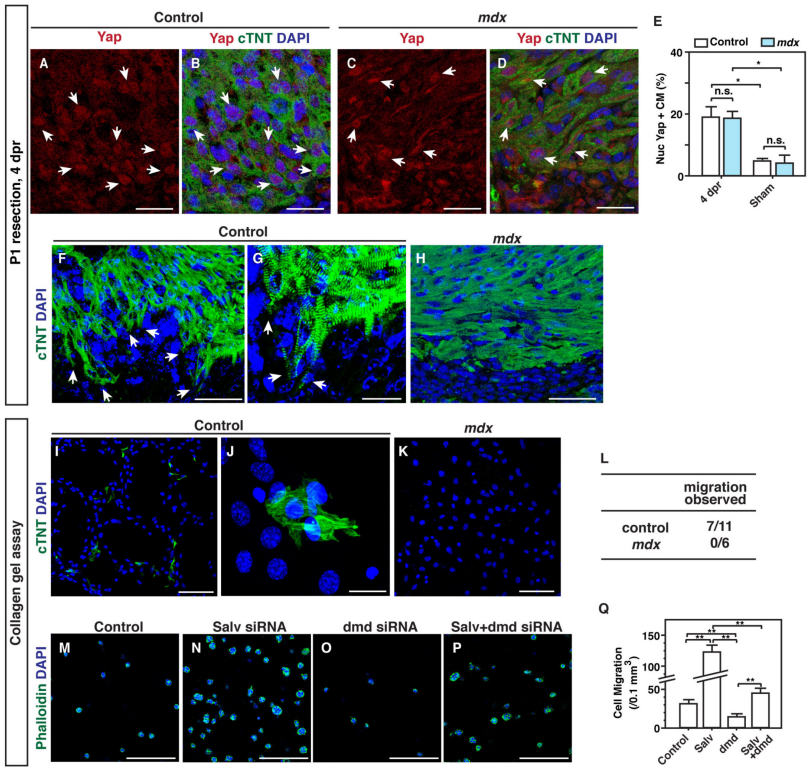


Figure 4.10: Regulation of cardiomyocyte protrusion by the dystrophin complex. (A-D) Yap localization in border zone cardiomyocytes of B6/10 control (A, B) and *Mdx*-B6/10 (C, D) mouse hearts at 4 dpr. Arrows show nuclear localized Yap. Bars = 25 µm. (E) Quantification of nuclear Yap in border zone cardiomyocytes at 4 dpr (n = 3 per genotype). *$p < 0.05$. (F-H) DGC is required for cardiomyocyte protrusion. Sections of the border zone of B6/10 control (F, G) and *Mdx*-B6/10 (H) mouse hearts were stained with the cardiac marker cTnt at 4 dpr. Arrows show protruding cells at the border zone. Panel G is a higher magnification image of F. For F and G, bars = 50 µm; for G, bars=25 µm. (I-K) Collagen gel assay for P1 B10 (I, J) and *Mdx*-B10 (K) hearts. Panel J is a higher magnification image of I. For I and K, bars = 100 µm; for J, bars = 20 µm. (L) Quantification of the migration observed in B10 control (n = 11) and *Mdx*-B10 (n = 6) mouse cardiomyocytes. p = 0.035, control *vs. Mdx*. (M-P) P19 cell migration in collagen gel after labeled siRNA treatment. Negative control (M), *Salv*(N), dystrophin (dmd) (O), and *Salv*and dmd combined (P). Bars = 250 µm. (Q) Quantification of migrated cells after each treatment. n = 3 for all groups. * $p < 0.01$. Figure credit: John Leach, Yuka Morikawa, Todd Heallen

subcellular localization is regulated by matrix rigidity, support the idea that Yap interprets and responds to local ECM mechanical characteristics [6, 31]. Yap-regulated genes are components of 2 plasmalemmal complexes that link the actin cytoskeleton to the ECM: the DGC and the talin-vinculin-integrin complex [3]. These genes include the DGC component *Sgcd* and the actin-binding protein *Talin2*, which links the actin cytoskeleton to integrins and ECM. Other Yap-regulated genes such as *Ctnna3*, a cadherin-associated gene localized to the intercalated disc (ICD), connect the actin cytoskeleton to both the ICD and ECM and most likely function to sense tension between cardiomyocytes [72]. Linking complexes are critical for cardiac muscle homeostasis. The DGC is disrupted in multiple types of muscular dystrophy with dilated cardiomyopathy (DCM). Mutations in Vinculin also cause DCM in humans [86]. Likewise, loss of function mutations in *Ctnna3* result in DCM, revealing a critical requirement of this protein in cardiomyocyte homeostasis [72]. Our molecular findings suggesting that the myocardium is linked to the ECM more efficiently in *Salv*CKO mouse hearts are also supported by the previously reported observation that *Lats1/2* mutant myocardium can transmit tissue deformations more effectively, as determined by using optical coherence tomography [140].

4.2.2 Actin cytoskeleton remodeling with cellular protrusion is required for heart regeneration

In the border zone of *Salv*CKO mouse hearts, we observed changes in cellular phenotypes, including the appearance of sarcomere-filled protrusions surrounding the wound in apex resection and adult myocardial infarction models. Our data also showed that *Salv*CKO mouse cardiomyocytes have the ability to invade a collagen gel. Protrusive activity was observed in both apex resection and adult LADO contexts. Our genomic analyses indicated that Yap regulates genes promoting lamellipodia forma-

tion and actin cytoskeleton protrusion. Certain Yap target genes encode proteins that localize to lamellipodia and filopodia such as *Enah*, an Ena/VASP actin regulator that causes cardiac dysfunction when disrupted in mice [87]. *Fgd4*, another Yap target gene, is mutated in Charcot-Marie-Tooth type 4H disease. *Fgd4* encodes an F-actin binding protein with RhoGEF activity specific for *Cdc42*, which is crucial to filopodia and lamellipodia formation and is essential for cytokinesis [141].

The Yap target gene *Mtss1*, also named missing in metastasis (MIM), encodes a protein that localizes to adherens junctions where it promotes actin filament assembly. *Mtss1* induces the formation of lamellipodial microspikes and membrane ruffles that contribute to protrusions and may be involved in the sensing of fluid shear stress in the kidney [114]. Other findings indicate that Mtss1 is required for neural tube closure [79]. Intriguingly, Yap is regulated by polymerized F actin, which promotes Yap nuclear localization, suggesting a positive feedback loop that maintains nuclear Yap localization and cellular protrusions.

4.2.3 Yap directly regulates genes encoding proteins that promote cell cycle progression and cytokinesis

Our findings indicated that apoptosis is unaffected in *Salv*CKO mice after myocardial infarction. Similar findings were previously reported for Yap gain-of-function mice [75]. Our Yap target gene dataset indicated that Yap directly regulates multiple genes that enhance progression through multiple cell cycle phases and cytokinesis. Yap directly activated the cytokinesis genes, *Aurkb* and *Birc5*. Another direct gene target of Yap, *Lin9*, encodes a *MuvB* complex component that enhances mitotic gene expression [115]. Interestingly, considerable cross-talk has been reported between the *MuvB* and *Lats2* in senescent cells, which warrants further investigation [131]. Other direct gene targets of Yap identified in our study included *Spin1*, a cytokinesis-associated gene ex-

pressed in cardiomyocytes, and *Aspm*, which is mutated in microcephaly patients and is required for proper spindle orientation and cytokinesis. The expression of *Aspm* is increased in cardiac hypertrophy, and it promotes Wnt signaling in neurogenesis [76].

Yap target genes identified in our study also included *Bub1b* and *Mad2l*, components of the spindle assembly checkpoint that ensure mitotic fidelity. *Bub1b* gain of function extends lifespan and inhibits aneuploidy and chromosome instability. In addition, other Yap target genes identified in this study were the cell cycle genes *Ccne2* and *Cdk6*, which play important roles in G1 progression. Moreover, the genes that encode 2 inhibitors of cell cycle progression, *Cdkn2b* and *Cdkn3*, were identified as Yap targets that likely function in a negative feedback mechanism. In some contexts, Cdkn3 may act to promote proliferation via poorly understood mechanism(s) [149].

4.2.4 Yap promotes the expression of its target genes that are more highly expressed in the fetal heart: evidence for dedifferentiation

Yap promoted the expression of its target genes in the postnatal heart, revealing a molecular mechanism for fetal gene program reversion that has been described but poorly understood mechanistically. Consistent with our molecular data, border zone cardiomyocytes of Hippo-deficient mouse hearts displayed a more primitive cardiomyocyte phenotype with more sarcomere disassembly than did border zone cardiomyocytes of control mouse hearts, suggesting that Hippo-deficient cardiomyocytes dedifferentiate after injury. Cardiomyocyte dedifferentiation also occurs in zebrafish heart regeneration, implying that there are retained mechanisms between fish and mammals [56]. The mechanisms for sarcomere disassembly in the context of regeneration are poorly understood but may involve regulated protein degradation.

4.3 Materials and methods

4.3.1 Mice

Mouse strains used in this study were as follows: *Myh6-CreErt; mTmG, Myh6-CreErt; Salvfx/fx; mTmG* [47]. C57BL/10J (B10), B10 crossed with C57BL/6J (B6/10), C57BL/10ScSn-Dmdmdx/J (*Mdx*-B10), and Mdx crossed with B6 (*Mdx*-B6/10). B10, B6, and *Mdx*-10 mice were obtained from Jackson Laboratory.

4.3.2 ChIP-seq analysis

Immunoprecipitation was performed with an anti-Yap antibody (Novus) in whole-heart lysates to generate libraries that were then sequenced by using an Ion Proton sequencer (Life Technologies). Two biological replicates were performed, and the reproducibility between samples was assessed by using Pearson's correlation coefficient. About 24 million Yap binding reads were evaluated by using a Homer software package.

4.3.3 Microarray and ChIP-seq analysis

Microarray analysis was performed on a Phalanx Mouse OneArray v2 platform. Two biological replicates, each with 3 technical replicates, were collected for each genotype. The value distribution of raw intensity from a total of 4 samples was graphically viewed in box plots. Linear correlation coefficients between biological replicates were measured by using R software function var.

Differentially expressed genes were detected by using the R bioconductor package *Limma*, version 3.16.8. Briefly, log transformation was performed to the raw intensity values; a Student t test was performed to identify differences in mRNA expression level between control and Salv CKO mouse hearts for each probe. Differentially expressed genes were called by using the cutoff p-value ≤ 0.05 and fold change ≥ 1.5. Activated

Yap1 targets were extracted from dataset GSE33019 by using the same criteria.

Ion Torrent Proton reads were mapped to the mm9 assembly (NCBI Build 37) by using Torrent Suite (2.0.1) aligner *Tmap* (0.2.3) (Life Technologies). Two biological replicates were performed. Only uniquely mapped reads were kept. A total of 11 million Yap ChIP-Seq reads were generated from replicate 1 and 13 million from replicate 2. The reproducibility between the 2 runs was measured by using Pearson's correlation coefficient to compare signal intensities in 5-kbp bins across the genome (Fig. S1G, r =0.69). ChIP-enriched peaks were identified from combined reads by using Homer v4.2 parameters *findPeaks -style factor -o -i input -tagThreshold 15*. The peak size was calculated automatically as 162 bp. The cutoffs for calling the peaks were read number enriched over input control values ≤ 4 fold, minimum read number of 15 which limited Poisson p-value to less than 1e-10. The clonal peaks were filtered out. This process produced 35,412 enriched peaks. Nearest genes associated to the enriched peaks were annotated by using the "annotatePeaks" function, and de novo motif discovery was performed by using the *findMotifsGenome* function in Homer.

A total of 1,706 genes had increased expression in *Salv*CKO mouse hearts and were overlaid with 10,396 Yap ChIP-Seq binding genes. Overlaid genes were subjected to over-represented gene ontology analysis by using GO Elite with 5000 permutations. The false discovery rate was calculated by using the Benjamini-Hochberg correction. Gene ontology terms were plotted by using the Z score. Only terms with a false discovery rate <0.10 were reported. Microarray and ChIP-Seq data are available through the NCBI Gene Expression Omnibus (GEO) data repository under accession GSE44103.

4.3.4 Human heart RNA-seq data analysis

Human RNA-Seq data were obtained from Gene Expression Omnibus (GEO) and DNA Data Bank of Japan (DDBJ). The sample IDs of fetal hearts were SRR643778,

SRR643779; of adult hearts: GSM1010964, ERR030894. The reads were aligned to reference genome hg19 by using Tophat. The splicing junctions were detected by using known annotations. Uniquely mapped reads within exons were counted with HTSeq-count. Differentially expressed genes were detected by using R package *DESeq* with a threshold false discovery rate (FDR)≤0.1. Gene ontology analysis of 700 genes with increased expression and 419 genes with reduced expression was performed in GO Elite.

4.3.5 Human DHS-seq data analysis

Mapped human DHS-Seq data were extracted from GEO. We used data for 11 fetal heart samples (GSM530654, GSM530661, GSM665811, GSM665814, GSM665817, GSM665824, GSM665830, GSM665831, GSM774203, GSM817220, GSM878630, bed file) and 2 adult heart samples (GSM1008559, bam file); tissue-specific motif calling was performed on a mouse DNaseI digital genomic footprinting dataset GSE40869. Peaks enriched in fetal hearts were detected against adult DHS data by using Homer with the following script: *findPeaks fetal_dhs -style dnase -o auto -i adult_dhs*. A total of 73,949 peaks were called. De novo motif enrichment was performed with the Homer function *findMotifsGenome*. Hierarchical clustering of enriched motifs in each tissue was computed and visualized on the basis of -log(*p*-value) by using R package pheatmap.

4.3.6 Mouse heart H3K27Ac ChIP-seq data analysis

Mouse heart H3K27Ac ChIP-Seq data were available in GEO (GSE52386). Bed-Graph files were generated from mapped bam files by using the Homer *makeUCSCfile* function and were visualized in the UCSC genome browser. The total read number of individual samples was normalized to 10 million. P7 H3K27Ac ChIP-Seq data were used to plot genome-wide density profiling within a 6-kb range of Yap binding peaks.

4.3.7 Left anterior descending coronary artery occlusion and apex resection

Left anterior descending coronary artery occlusion (LADO) was performed as previously described on P8 or adult (8-10 week old) mice (8). In the postnatal (P8) 1 dpmi mouse study, tamoxifen (0.5 mg) was injected in *Myh6-CreErt; mTmG* (control) and *Myh6-CreErt; Salvfx/fx; mTmG* (*Salv*CKO) mice 1 day before and 5 min after surgery. For 8-10 week old mice, tamoxifen (1.5 mg) was injected 7 and 6 days before and within 2 hours (2mg) after surgery. Hearts were collected at 4, 7, and 15 days after LADO, and EdU was injected 4 times starting at 3 days before collection of the heart.

Resection of the heart apex was performed on P1 mice as previously described (8) by using B6/10 or B10 (control) and *Mdx*(with B6/10 or B10 background) mice. Hearts were collected at either 4 or 21 days after resection. At 21 days after resection before collection, echocardiography was performed in resected and sham animals. All echocardiographic measurements were performed in the Mouse Phenotyping Core at Baylor College of Medicine. One-way ANOVA with Bonferroni's multiple comparison test was performed for statistical analysis.

After echocardiography, hearts were collected, embedded in paraffin, sectioned (7-μm thickness), and stained by using Masson's trichrome staining. After images were documented, scar size was measured by using ImageJ. Every other 3 slides were documented, and the sizes of the sections with the largest scars were recorded. The Mann-Whitney U test was performed for statistical analysis.

4.3.8 Edu staining, TUNEL assay, and immunohistochemistry

For the adult mouse study, EdU (5'ethynyl-2'-deoxyuridine) was injected into mice for 4 consecutive days starting 3 days before the heart was harvested. For the P8 mouse study, hearts were collected at 1 day post myocardial infacrtion, and EdU was injected within 5 min after surgery and at 4 h before the heart was harvested. For the

neonatal mouse study, hearts were collected at 4 days post resection, and EdU was injected 4 h before tissue collection. Hearts were embedded in paraffin and sectioned (7- μm thickness). EdU staining was performed by using the Click-iT EdU Alexa Fluor 647 Imaging kit (Life Technologies). Immunohistochemistry with eGFP or cTNT was performed to label cardiomyocytes after EdU staining.

The TUNEL assay was performed by using the DeadEnd fluorometric TUNEL assay (Promega). To mark cardiomyocytes, immunohistochemistry with cTNT was performed after TUNEL staining.

Immunohistochemistry was performed by using standard protocols. Antibodies used in this study were as follows: rabbit anti-GFP (1:400 dilution, Abcam), rabbit anti-AuroraB kinase (1:200 dilution, Abcam), mouse anti-GFP (1:500 dilution, Clontech), rabbit anti-RFP (1:200 dilution, Clontech), rabbit anti-Yap (1:200 dilution, Novus), mouse anti-cTNT (1:200 dilution, Thermo), rabbit anti-talin (1:200 dilution, Abcam), and mouse anti-vinculin (1:200 dilution, Thermo). Fluorescence images were documented by using a Leica TCS SP5 confocal microscope with Leica LAS AF software. All images presented in the manuscript were prepared by using Adobe Photoshop CS5.1 (Adobe Systems). Border zone cardiomyocytes were defined as cells within 300 μm from the scar.

For the quantification of positive cells, 3 biological replicates were collected per genotype and time points. For each stain, 3 images were captured for each biological replicate. Cell counting was performed blindly. A total of 100-250 cardiomyocytes were counted in each image for quantification. For the quantification of Aurora B kinase, a total of 10 sections were counted per biological replicate. For quantitation of protrusions, 100 cells adjacent to the scar were examined for the number of protrusions (less than 10 μm, 10-20 μm, and greater than 20 μm). The Mann-Whitney U test was performed for statistical analysis.

4.3.9 Collagen gel assay with heart tissue

For the study of P10 hearts, *Myh6-CreErt; mTmG* (control) and *Myh6-CreErt; Salvfx/fx; mTmG* (*Salv*CKO) mice were used. Tamoxifen (0.6 mg) was injected twice at P8 and P9, and hearts were dissected at P10. Apexes were cut and placed on collagen gel (Millipore). Tissues were mounted by overlaying collagen gel and were cultured in DMEM with 10% bovine calf serum (BCS) at 37°C. After 5 days, collagen gels were collected and stained with eGFP. Each gel was analyzed for the presence of cardiomyocytes that migrated to the bottom gel. If any cardiomyocytes were found in the bottom gel, it was counted as "migration observed." The Fisher's exact test was performed for statistical analysis.

For the study of P1 hearts, B10 (control) and *Mdx*-B10 mice were used. Hearts were harvested at P1, and apexes were collected. Nylon mesh (100 μm) was placed on the collagen gel, and the tissues were placed on top of the mesh. Tissues were mounted by overlaying collagen gel and were cultured in DMEM media with 10% BCS at 37˚C. After 5 days, the top gel and nylon mesh were removed, and the bottom collagen gel was stained with cTNT. Quantification was performed as described above.

4.3.10 Luciferase assay and transfection experiments

Each enhancer element was amplified and cloned into the pGL3-promoter plasmid (Promega). The primers used are listed in Supplemental Table 1. The Tead motif for each element was deleted by using the Quick Change XL Site-directed mutagenesis kit (Agilent Technologies) for *Sgcd* or Gibson assembly kit (NEB) for *Aurkb*, *Ctnna3*, *Enah*, *Fmn2*, and *Lin9*. The locations and sequences of TEAD motifs are listed in Table S1. Plasmids were co-transfected with Yap and Tead expression plasmids (7) into P19 embryonal carcinoma cells. Luciferase activity was analyzed by using the Dual Glo luciferase assay system (Promega). Three independent transfection experiments were

performed with triplicate wells. Mann-Whitney test was used for statistical analysis.

4.3.11 P19 migration assay

To examine the influence of Yap target genes on cell motility, we performed siRNA experiments in P19 embryonal carcinoma cells. siRNA oligos were obtained from Integrated DNA Technologies (Coralville, Iowa). The methods used for siRNA knockdown experiments have been described previously (8). After 24 hours of siRNA treatment, trypsinized cells were suspended in collagen gel mix (Millipore). Nylon mesh (20 μm) was applied to solidified collagen gel spots in 4-well plates, and cell/collagen mixes were dispensed to meshes. Plates were incubated at 37°C (5% CO2) for 1 h and α-MEM culture media with 10% fetal bovine serum (FBS) was subsequently added. After 24 h of culture, the top gel and nylon mesh were removed, and the bottom gel was immmunostained accordingly. We examined the statistical significance of the differences between groups by using Mann-Whitney U tests. $p < 0.05$ was considered significant.

4.3.12 Western blotting

Methods for Western Blot quantitation were previously described (7). Primary antibodies used were anti-α-tubulin (1:8000 dilution; Sigma) and anti-SGCD (1:500 dilution; Santa Cruz Biotechnology). Differences between groups were evaluated for statistical significance by using two-way ANOVA with the post-hoc Bonferroni correction.

4.3.13 Quantitative PCR

Total RNA was extracted from dissected heart tissues by using the RNeasy Micro Kit (Qiagen), and the concentration was determined by using an Infinite M200PRO Nanodrop spectrophotometer (Tecan). First-strand cDNA synthesis was performed by

using the iScript cDNA synthesis kit (Invitrogen) using 1 μg of total RNA input. For qRT-PCR, 1 μl of cDNA (1:10 dilution) was added to SYBR Green Jumpstart Taq Ready Mix (Sigma). Each sample was analyzed in triplicate and run on a StepOnePlus Real-Time PCR System (Life Technologies). Mean dCt values for each Yap target gene were normalized against those of GAPDH mRNA levels, and corresponding ddCt values were log2-transformed to obtain fold-change values. The standard deviations of control and mutant samples were calculated within groups. The non-parametric Mann-Whitney test was performed to determine statistical significance of the differences between control and mutant groups for each target gene ($p < 0.05$ was considered significant). Sequences of primers used in qPCR experiments are available upon request.

4.3.14 Cell surface area measurements

Immunofluorescence staining for GFP and DAPI was performed as described previously [47]. For visualization of cell membranes, samples were incubated with wheat germ agglutin Alexa Fluor 647 conjugate (1:200 dilution; Life Technologies) for 30 min at room temperature before blocking and primary antibody application. Cardiomyocyte cell surface area measurements were obtained from multiple transverse sections in series, and averages were obtained by using ImageJ software. For 4 and 10 dpmi hearts, we used *Mhccre-Ert; mTmG* (control) and *Mhccre-Ert; mTmG; Salvf/f* (*Salv*CKO) mice. For 21 dpmi and 21 dpmi sham hearts, we used *Mhccre-Ert* (control) and *Mhccre-Ert; Salvf/f* (*Salv*CKO) mice. Differences between groups were examined for statistical significance by using The Mann-Whitney U tests. $p < 0.05$ was considered significant.

4.3.15 4C sequencing and analysis

The 4C templates were generated from P8 *Salv*CKO mouse hearts (2 biological replicates), as previously described [124, 125]. HindIII and DpnII were used as first and second restriction enzymes, respectively. Promoter sequences from *Ctnna3* and *Sgcd* were used to design viewpoint amplification primers (Ctnna3 DpnII: 5' ACTGGCTGTGGAGGTTTTGAATTAC; HindIII: 5' CCCTCTGTATTGACTTTCTATCCC; Sgcd DpnII: 5' GAAACCCTACTCTGCCTCTACTTC; HindIII: 5' CAGAGGGACTGGGTGCTTCT). Libraries were amplified from 100 ng of template DNA by using the following PCR protocol: 94°C 2 min, 30 cycles of 94°C 15 s, 56 °C 1 min, 68 °C 1 min. The libraries were then prepared for Ion Proton Sequencer. Ion Proton reads were separated by PCR primers. Over 700,000 reads were generated for each viewpoint library. For extracting unknown contact point sequences, only reads were kept that contained a recognition restriction enzyme site jointed to the PCR primer and, at most, one secondary restriction site. After trimming the primer sequences, unknown sequences with length ≥ 40 bp were kept for aligning against mouse genome mm9. Significant contact points in the proximal range (500 kb) were detected by using a previously described algorithm [124, 120] and were visualized in the UCSC genome browser.

5. SUMMARY AND CONCLUSIONS

My dissertation work has focused on using bioinformatic analysis to identify gene regulatory programs that directly involved in atrial fibrillation and heart regeneration.

Pitx2, a transcription factor that controls transcription of numerous target genes, is located in proximity to sequence variants in the human genome that are commonly associated with atrial fibrillation (AF). While previous work focused on *Pitx2* during development, the current study investigated *Pitx2* in postnatal heart. It was previously unknown whether Pitx2 has distinct postnatal and developmental functions. The *Pitx2* genes was removed from hearts of postnatal mice. Unstressed adult Pitx2 mutant mice display variable R-R interval with diminished P-wave amplitude characteristic of arrhythmias called sinus node dysfunction, an AF risk factor in human patients. To uncover target genes that are regulated by Pitx2, we used a genome-wide approach, overlaid differentially expressed genes in *Pitx2* mutants as well as genes directly bound by Pitx2, and identify direct targets of Pitx2. Pitx2 target genes encoded cell junction proteins, ion channels, and critical transcriptional regulators. Importantly, many Pitx2 target genes have been previously implicated in human AF. Other studies in adult *Pitx2* mutant mice revealed structural remodeling in the heart characteristic of human AF patients. Our findings provide new mechanistic insight into AF, revealing that *Pitx2* has genetically separable postnatal and developmental functions and unveiling direct Pitx2 target genes that include channel and calcium handling genes as well as genes that stabilize the cellular structure in postnatal atrium. Since Pitx2 regulates many of these genes in the postnatal heart, it is conceivable that drugs can be developed to modulate the molecular interaction between Pitx2 and its target genes. Additionally, we performed chromosome conformation capture combined with high-

throughput sequencing (4C-Seq) using *Pitx2* promoters and regulatory regions overlaid with atrial fibrillation high-risk locus. Our results indicate direct contacts of *Pitx2* promoter and AF-associated enhancer region. We deleted the 20-kb region *in vitro*, and found the up-stream enhancer directly regulates *Pitx2ab* isoform, not *Pitx2c*. Such distinct enhancer-promoter specificity is mediated by genome insulator protein CTCF. Depletion of CTCF site that segregates *Pitx2ab* and *Pitx2c* promoters led to restore of *Pitx2ab* expression. Our findings that many important *Pitx2* regulated events occur postnatally also strengthens the likelihood that *Pitx2*-mediated atrial fibrillation may be treatable in the future.

We also identified key components that are essential for heart regeneration in hippo-deficient mice. Hippo signaling is an evolutionarily conserved kinase cascade that regulates organ size during development and prevents adult mammalian cardiomyocyte regeneration by inhibiting the transcriptional coactivator Yap, which also responds to mechanical signaling in cultured cells to promote cell proliferation. To identify Yap target genes that are activated during cardiomyocyte renewal and regeneration, we performed Yap chromatin immunoprecipitation sequencing (ChIP-Seq) and mRNA expression profiling in Hippo signaling–deficient mouse hearts. We found that Yap directly regulated genes encoding cell cycle progression proteins, as well as genes encoding proteins that promote F-actin polymerization and that link the actin cytoskeleton to the extracellular matrix. Included in the latter group were components of the dystrophin glycoprotein complex, a large molecular complex that, when defective, results in muscular dystrophy in humans. Cardiomyocytes near the scar tissue of injured Hippo signaling—deficient mouse hearts showed cellular protrusions suggestive of cytoskeletal remodeling. The hearts of mdx mutant mice, which lack functional dystrophin and are a model for muscular dystrophy, showed impaired regeneration and cytoskeleton remodeling, but normal cardiomyocyte proliferation, after injury. Our

data showed that, in addition to genes encoding cell cycle progression proteins, Yap regulated genes that enhance cytoskeletal remodeling. Thus, blocking the Hippo pathway input to Yap may tip the balance so that Yap responds to mechanical changes associated with heart injury to promote repair.

In conclusion, integrated bioinformatic analysis allows discovery of novel gene regulation programs. Through large-scale analysis, we were able to dissect direct targets of master regulators like *Pitx2* in arrhythmias and *Yap* in cardiac regeneration. The philosophy behind the bioinformatic analysis is to extract the similarity between different systems, such as human and mouse genome, or gene programs involved in development and regeneration, thus build connection between elements of each system. The ultimate goal of bioinformatic analysis is not to abandon bench works but to narrow down the list of candidate genes and regulatory elements and to facilitate strong hypothesis for future experimental test. My future work will focus on mathematical models with machine learning to understand the effect of genomic variants in cardiac development and diseases.

REFERENCES

[1] B. A. Amendt, L. B. Sutherland, E. V. Semina, and A. F. Russo. The molecular basis of rieger syndrome. analysis of pitx2 homeodomain protein activities. *J Biol Chem*, 273(32):20066–72, Aug 1998.

[2] G. Ammirabile, A. Tessari, V. Pignataro, D. Szumska, F. Sutera Sardo, J. Benes, Jr, M. Balistreri, S. Bhattacharya, D. Sedmera, and M. Campione. Pitx2 confers left morphological, molecular, and functional identity to the sinus venosus myocardium. *Cardiovasc Res*, 93(2):291–301, Feb 2012.

[3] G. Anastasi, G. Cutroneo, R. Gaeta, D. Di Mauro, A. Arco, A. Consolo, G. Santoro, F. Trimarchi, and A. Favaloro. Dystrophin-glycoprotein complex and vinculin-talin-integrin system in human adult cardiac muscle. *Int J Mol Med*, 23(2):149–59, Feb 2009.

[4] S. Anders, P. T. Pyl, and W. Huber. Htseq–a python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–9, Jan 2015.

[5] B. D. Angst, L. U. Khan, N. J. Severs, K. Whitely, S. Rothery, R. P. Thompson, A. I. Magee, and R. G. Gourdie. Dissociated spatial patterning of gap junctions and cell adhesion junctions during postnatal differentiation of ventricular myocardium. *Circ Res*, 80(1):88–94, Jan 1997.

[6] M. Aragona, T. Panciera, A. Manfrin, S. Giulitti, F. Michielin, N. Elvassore, S. Dupont, and S. Piccolo. A mechanical checkpoint controls multicellular growth through yap/taz regulation by actin-processing factors. *Cell*, 154(5):1047–59, Aug 2013.

[7] S. M. Baig, A. Koschak, A. Lieb, M. Gebhart, C. Dafinger, G. Nürnberg, A. Ali, I. Ahmad, M. J. Sinnegger-Brauns, N. Brandt, J. Engel, M. E. Mangoni, M. Farooq, H. U. Khan, P. Nürnberg, J. Striessnig, and H. J. Bolz. Loss of ca(v)1.3 (cacna1d) function in a human channelopathy with bradycardia and congenital deafness. *Nat Neurosci*, 14(1):77–84, Jan 2011.

[8] E. Balse, D. F. Steele, H. Abriel, A. Coulombe, D. Fedida, and S. N. Hatem. Dynamic of ion channel expression at the plasma membrane of cardiomyocytes. *Physiol Rev*, 92(3):1317–58, Jul 2012.

[9] E. R. Barton. Impact of sarcoglycan complex on mechanical signal transduction in murine skeletal muscle. *Am J Physiol Cell Physiol*, 290(2):C411–9, Feb 2006.

[10] E. J. Benjamin, P.-S. Chen, D. E. Bild, A. M. Mascette, C. M. Albert, A. Alonso, H. Calkins, S. J. Connolly, A. B. Curtis, D. Darbar, P. T. Ellinor, A. S. Go, N. F. Goldschlager, S. R. Heckbert, J. Jalife, C. R. Kerr, D. Levy, D. M. Lloyd-Jones, B. M. Massie, S. Nattel, J. E. Olgin, D. L. Packer, S. S. Po, T. S. M. Tsang, D. R. Van Wagoner, A. L. Waldo, and D. G. Wyse. Prevention of atrial fibrillation: report from a national heart, lung, and blood institute workshop. *Circulation*, 119(4):606–18, Feb 2009.

[11] O. Bergmann, R. D. Bhardwaj, S. Bernard, S. Zdunek, F. Barnabé-Heider, S. Walsh, J. Zupicich, K. Alkass, B. A. Buchholz, H. Druid, S. Jovinge, and J. Frisén. Evidence for cardiomyocyte renewal in humans. *Science*, 324(5923):98–102, Apr 2009.

[12] B. E. Bernstein, J. A. Stamatoyannopoulos, J. F. Costello, B. Ren, A. Milosavljevic, A. Meissner, M. Kellis, M. A. Marra, A. L. Beaudet, J. R. Ecker, P. J. Farnham, M. Hirst, E. S. Lander, T. S. Mikkelsen, and J. A. Thomson. The nih

roadmap epigenomics mapping consortium. *Nat Biotechnol*, 28(10):1045–8, Oct 2010.

[13] L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6):947–56, Sep 2005.

[14] J. C. Brüning, M. D. Michael, J. N. Winnay, T. Hayashi, D. Hörsch, D. Accili, L. J. Goodyear, and C. R. Kahn. A muscle-specific insulin receptor knockout exhibits features of the metabolic syndrome of niddm without altering glucose tolerance. *Mol Cell*, 2(5):559–69, Nov 1998.

[15] M. G. Chelu, S. Sarma, S. Sood, S. Wang, R. J. van Oort, D. G. Skapura, N. Li, M. Santonastasi, F. U. Müller, W. Schmitz, U. Schotten, M. E. Anderson, M. Valderrábano, D. Dobrev, and X. H. T. Wehrens. Calmodulin kinase ii-mediated sarcoplasmic reticulum ca2+ leak promotes atrial fibrillation in mice. *J Clin Invest*, 119(7):1940–51, Jul 2009.

[16] Y.-H. Chen, S.-J. Xu, S. Bendahhou, X.-L. Wang, Y. Wang, W.-Y. Xu, H.-W. Jin, H. Sun, X.-Y. Su, Q.-N. Zhuang, Y.-Q. Yang, Y.-B. Li, Y. Liu, H.-J. Xu, X.-F. Li, N. Ma, C.-P. Mou, Z. Chen, J. Barhanin, and W. Huang. Kcnq1 gain-of-function mutation in familial atrial fibrillation. *Science*, 299(5604):251–4, Jan 2003.

[17] S. Cheng, M. J. Keyes, M. G. Larson, E. L. McCabe, C. Newton-Cheh, D. Levy, E. J. Benjamin, R. S. Vasan, and T. J. Wang. Long-term outcomes in individuals with prolonged pr interval or first-degree atrioventricular block. *JAMA*, 301(24):2571–7, Jun 2009.

[18] D. Y. Chiang, M. Zhang, N. Voigt, K. M. Alsina, H. Jakob, J. F. Martin, D. Dobrev, X. H. T. Wehrens, and N. Li. Identification of microrna-mrna dysregulations in paroxysmal atrial fibrillation. *Int J Cardiol*, 184:190–7, Apr 2015.

[19] A. Chinchilla, H. Daimi, E. Lozano-Velasco, J. N. Dominguez, R. Caballero, E. Delpón, J. Tamargo, J. Cinca, L. Hove-Madsen, A. E. Aranega, and D. Franco. Pitx2 insufficiency leads to atrial electrical and structural remodeling linked to arrhythmogenesis. *Circ Cardiovasc Genet*, 4(3):269–79, Jun 2011.

[20] J. J. H. Chong, V. Chandrakanthan, M. Xaymardan, N. S. Asli, J. Li, I. Ahmed, C. Heffernan, M. K. Menon, C. J. Scarlett, A. Rashidianfar, C. Biben, H. Zoellner, E. K. Colvin, J. E. Pimanda, A. V. Biankin, B. Zhou, W. T. Pu, O. W. J. Prall, and R. P. Harvey. Adult cardiac-resident msc-like stem cells with a proepicardial origin. *Cell Stem Cell*, 9(6):527–40, Dec 2011.

[21] O. Corradin, A. Saiakhova, B. Akhtar-Zaidi, L. Myeroff, J. Willis, R. Cowper-Sal lari, M. Lupien, S. Markowitz, and P. C. Scacheri. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res*, 24(1):1–13, Jan 2014.

[22] T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2(4):292–301, Apr 2001.

[23] W. de Laat and D. Duboule. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, 502(7472):499–506, Oct 2013.

[24] E. de Wit and W. de Laat. A decade of 3c technologies: insights into nuclear organization. *Genes Dev*, 26(1):11–24, Jan 2012.

[25] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner. Capturing chromosome conformation. *Science*, 295(5558):1306–11, Feb 2002.

[26] W. P. Devine, J. D. Wythe, M. George, K. Koshiba-Takeuchi, and B. G. Bruneau. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *Elife*, 3, 2014.

[27] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–80, May 2012.

[28] D. Dobrev, N. Voigt, and X. H. T. Wehrens. The ryanodine receptor channel as a molecular motif in atrial fibrillation: pathophysiological and therapeutic implications. *Cardiovasc Res*, 89(4):734–43, Mar 2011.

[29] D. Dobrev and X. H. T. Wehrens. Calmodulin kinase ii, sarcoplasmic reticulum ca2+ leak, and atrial fibrillation. *Trends Cardiovasc Med*, 20(1):30–4, Jan 2010.

[30] J. M. Dowen, Z. P. Fan, D. Hnisz, G. Ren, B. J. Abraham, L. N. Zhang, A. S. Weintraub, J. Schuijers, T. I. Lee, K. Zhao, and R. A. Young. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, 159(2):374–87, Oct 2014.

[31] S. Dupont, L. Morsut, M. Aragona, E. Enzo, S. Giulitti, M. Cordenonsi, F. Zanconato, J. Le Digabel, M. Forcato, S. Bicciato, N. Elvassore, and S. Piccolo. Role of yap/taz in mechanotransduction. *Nature*, 474(7350):179–83, Jun 2011.

[32] P. T. Ellinor, K. L. Lunetta, C. M. Albert, N. L. Glazer, M. D. Ritchie, A. V. Smith, D. E. Arking, M. Müller-Nurasyid, B. P. Krijthe, S. A. Lubitz, J. C. Bis, M. K. Chung, M. Dörr, K. Ozaki, J. D. Roberts, J. G. Smith, A. Pfeufer, M. F. Sinner, K. Lohman, J. Ding, N. L. Smith, J. D. Smith, M. Rienstra, K. M. Rice,

D. R. Van Wagoner, J. W. Magnani, R. Wakili, S. Clauss, J. I. Rotter, G. Steinbeck, L. J. Launer, R. W. Davies, M. Borkovich, T. B. Harris, H. Lin, U. Völker, H. Völzke, D. J. Milan, A. Hofman, E. Boerwinkle, L. Y. Chen, E. Z. Soliman, B. F. Voight, G. Li, A. Chakravarti, M. Kubo, U. B. Tedrow, L. M. Rose, P. M. Ridker, D. Conen, T. Tsunoda, T. Furukawa, N. Sotoodehnia, S. Xu, N. Kamatani, D. Levy, Y. Nakamura, B. Parvez, S. Mahida, K. L. Furie, J. Rosand, R. Muhammad, B. M. Psaty, T. Meitinger, S. Perz, H.-E. Wichmann, J. C. M. Witteman, W. H. L. Kao, S. Kathiresan, D. M. Roden, A. G. Uitterlinden, F. Rivadeneira, B. McKnight, M. Sjögren, A. B. Newman, Y. Liu, M. H. Gollob, O. Melander, T. Tanaka, B. H. C. Stricker, S. B. Felix, A. Alonso, D. Darbar, J. Barnard, D. I. Chasman, S. R. Heckbert, E. J. Benjamin, V. Gudnason, and S. Kääb. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet*, 44(6):670–5, Jun 2012.

[33] ENCODE Project Consortium. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57–74, Sep 2012.

[34] T. H. Everett, 4th and J. E. Olgin. Atrial fibrosis and the mechanisms of atrial fibrillation. *Heart Rhythm*, 4(3 Suppl):S24–7, Mar 2007.

[35] X. Fang, W. Mei, W. B. Barbazuk, S. A. Rivkees, and C. C. Wendler. Caffeine exposure alters cardiac gene expression in embryonic cardiomyocytes. *Am J Physiol Regul Integr Comp Physiol*, 307(12):R1471–87, Dec 2014.

[36] M. J. Fullwood, M. H. Liu, Y. F. Pan, J. Liu, H. Xu, Y. B. Mohamed, Y. L. Orlov, S. Velkov, A. Ho, P. H. Mei, E. G. Y. Chew, P. Y. H. Huang, W.-J. Welboren, Y. Han, H. S. Ooi, P. N. Ariyaratne, V. B. Vega, Y. Luo, P. Y. Tan, P. Y. Choy, K. D. S. A. Wansa, B. Zhao, K. S. Lim, S. C. Leow, J. S. Yow, R. Joseph, H. Li, K. V. Desai, J. S. Thomsen, Y. K. Lee, R. K. M. Karuturi, T. Herve, G. Bourque,

H. G. Stunnenberg, X. Ruan, V. Cacheux-Rataboul, W.-K. Sung, E. T. Liu, C.-L. Wei, E. Cheung, and Y. Ruan. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, 462(7269):58–64, Nov 2009.

[37] Y. Ghavi-Helm, F. A. Klein, T. Pakozdi, L. Ciglar, D. Noordermeer, W. Huber, and E. E. M. Furlong. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*, 512(7512):96–100, Aug 2014.

[38] J. Giudice, Z. Xia, E. T. Wang, M. A. Scavuzzo, A. J. Ward, A. Kalsotra, W. Wang, X. H. T. Wehrens, C. B. Burge, W. Li, and T. A. Cooper. Alternative splicing regulates vesicular trafficking genes in cardiomyocytes during postnatal heart development. *Nat Commun*, 5:3603, 2014.

[39] A. Goyenvalle. Therapeutic approaches to muscular dystrophies. *Curr Gene Ther*, 12(3):137–8, Jun 2012.

[40] P. Grote, L. Wittler, D. Hendrix, F. Koch, S. Währisch, A. Beisaw, K. Macura, G. Bläss, M. Kellis, M. Werber, and B. G. Herrmann. The tissue-specific lncrna fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell*, 24(2):206–14, Jan 2013.

[41] D. F. Gudbjartsson, H. Holm, S. Gretarsdottir, G. Thorleifsson, G. B. Walters, G. Thorgeirsson, J. Gulcher, E. B. Mathiesen, I. Njølstad, A. Nyrnes, T. Wilsgaard, E. M. Hald, K. Hveem, C. Stoltenberg, G. Kucera, T. Stubblefield, S. Carter, D. Roden, M. C. Y. Ng, L. Baum, W. Y. So, K. S. Wong, J. C. N. Chan, C. Gieger, H.-E. Wichmann, A. Gschwendtner, M. Dichgans, G. Kuhlenbäumer, K. Berger, E. B. Ringelstein, S. Bevan, H. S. Markus, K. Kostulas, J. Hillert, S. Sveinbjörnsdóttir, E. M. Valdimarsson, M.-L. Løchen, R. C. W. Ma, D. Darbar, A. Kong, D. O. Arnar, U. Thorsteinsdottir, and K. Stefansson.

A sequence variant in zfhx3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet*, 41(8):876–8, Aug 2009.

[42] M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn, and E. S. Lander. Chromatin signature reveals over a thousand highly conserved large non-coding rnas in mammals. *Nature*, 458(7235):223–7, Mar 2009.

[43] M. Guttman, J. Donaghey, B. W. Carey, M. Garber, J. K. Grenier, G. Munson, G. Young, A. B. Lucas, R. Ach, L. Bruhn, X. Yang, I. Amit, A. Meissner, A. Regev, J. L. Rinn, D. E. Root, and E. S. Lander. lincrnas act in the circuitry controlling pluripotency and differentiation. *Nature*, 477(7364):295–300, Sep 2011.

[44] M. Haïssaguerre, P. Jaïs, D. C. Shah, A. Takahashi, M. Hocini, G. Quiniou, S. Garrigue, A. Le Mouroux, P. Le Métayer, and J. Clémenty. Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *N Engl J Med*, 339(10):659–66, Sep 1998.

[45] P. Han, W. Li, C.-H. Lin, J. Yang, C. Shang, S. T. Nurnberg, K. K. Jin, W. Xu, C.-Y. Lin, C.-J. Lin, Y. Xiong, H.-C. Chien, B. Zhou, E. Ashley, D. Bernstein, P.-S. Chen, H.-S. V. Chen, T. Quertermous, and C.-P. Chang. A long noncoding rna protects the heart from pathological hypertrophy. *Nature*, 514(7520):102–6, Oct 2014.

[46] T. Hashimshony, F. Wagner, N. Sher, and I. Yanai. Cel-seq: single-cell rna-seq by multiplexed linear amplification. *Cell Rep*, 2(3):666–73, Sep 2012.

[47] T. Heallen, Y. Morikawa, J. Leach, G. Tao, J. T. Willerson, R. L. Johnson, and J. F. Martin. Hippo signaling impedes adult heart regeneration. *Development*,

140(23):4683–90, Dec 2013.

[48] T. Heallen, M. Zhang, J. Wang, M. Bonilla-Claudio, E. Klysik, R. L. Johnson, and J. F. Martin. Hippo pathway inhibits wnt signaling to restrain cardiomyocyte proliferation and heart size. *Science*, 332(6028):458–61, Apr 2011.

[49] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol Cell*, 38(4):576–89, May 2010.

[50] H. Holm, D. F. Gudbjartsson, D. O. Arnar, G. Thorleifsson, G. Thorgeirsson, H. Stefansdottir, S. A. Gudjonsson, A. Jonasdottir, E. B. Mathiesen, I. Njølstad, A. Nyrnes, T. Wilsgaard, E. M. Hald, K. Hveem, C. Stoltenberg, M.-L. Løchen, A. Kong, U. Thorsteinsdottir, and K. Stefansson. Several common variants modulate heart rate, pr interval and qrs duration. *Nat Genet*, 42(2):117–22, Feb 2010.

[51] J. C.-G. Hombría and B. Lovegrove. Beyond homeosis–hox function in morphogenesis and organogenesis. *Differentiation*, 71(8):461–76, Oct 2003.

[52] J. Hsu, P. Hanna, D. R. Van Wagoner, J. Barnard, D. Serre, M. K. Chung, and J. D. Smith. Whole genome expression differences in human left and right atria ascertained by rna sequencing. *Circ Cardiovasc Genet*, 5(3):327–35, Jun 2012.

[53] D. Husser, V. Adams, C. Piorkowski, G. Hindricks, and A. Bollmann. Chromosome 4q25 variants and atrial fibrillation recurrence after catheter ablation. *J Am Coll Cardiol*, 55(8):747–53, Feb 2010.

[54] J. F. Islas, Y. Liu, K.-C. Weng, M. J. Robertson, S. Zhang, A. Prejusa, J. Harger, D. Tikhomirova, M. Chopra, D. Iyer, M. Mercola, R. G. Oshima, J. T. Willer-

son, V. N. Potaman, and R. J. Schwartz. Transcription factors ets2 and mesp1 transdifferentiate human dermal fibroblasts into cardiac progenitors. *Proc Natl Acad Sci U S A*, 109(32):13016–21, Aug 2012.

[55] D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay, and I. Amit. Massively parallel single-cell rna-seq for marker-free decomposition of tissues into cell types. *Science*, 343(6172):776–9, Feb 2014.

[56] C. Jopling, E. Sleep, M. Raya, M. Martí, A. Raya, and J. C. Izpisúa Belmonte. Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature*, 464(7288):606–9, Mar 2010.

[57] S. Kääb, D. Darbar, C. van Noord, J. Dupuis, A. Pfeufer, C. Newton-Cheh, R. Schnabel, S. Makino, M. F. Sinner, P. J. Kannankeril, B. M. Beckmann, S. Choudry, B. S. Donahue, J. Heeringa, S. Perz, K. L. Lunetta, M. G. Larson, D. Levy, C. A. MacRae, J. N. Ruskin, A. Wacker, A. Schömig, H.-E. Wichmann, G. Steinbeck, T. Meitinger, A. G. Uitterlinden, J. C. M. Witteman, D. M. Roden, E. J. Benjamin, and P. T. Ellinor. Large scale replication and meta-analysis of variants on chromosome 4q25 associated with atrial fibrillation. *Eur Heart J*, 30(7):813–9, Apr 2009.

[58] D. G. Katritsis, E. Giazitzoglou, S. Korovesis, E. Karvouni, C. E. Anagnostopoulos, and A. J. Camm. Conduction patterns in the cardiac veins: electrophysiologic characteristics of the connections between left atrial and coronary sinus musculature. *J Interv Card Electrophysiol*, 10(1):51–8, Feb 2004.

[59] S. J. Kattman, A. D. Witty, M. Gagliardi, N. C. Dubois, M. Niapour, A. Hotta, J. Ellis, and G. Keller. Stage-specific optimization of activin/nodal and bmp

signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell*, 8(2):228–40, Feb 2011.

[60] S. Kharche, C. J. Garratt, M. R. Boyett, S. Inada, A. V. Holden, J. C. Hancox, and H. Zhang. Atrial proarrhythmia due to increased inward rectifier current (i(k1)) arising from kcnj2 mutation–a simulation study. *Prog Biophys Mol Biol*, 98(2-3):186–97, 2008.

[61] K. Kikuchi and K. D. Poss. Cardiac regenerative capacity and mechanisms. *Annu Rev Cell Dev Biol*, 28:719–41, 2012.

[62] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. Tophat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 14(4):R36, 2013.

[63] P. Kirchhof, P. C. Kahr, S. Kaese, I. Piccini, I. Vokshi, H.-H. Scheld, H. Rotering, L. Fortmueller, S. Laakmann, S. Verheule, U. Schotten, L. Fabritz, and N. A. Brown. Pitx2c is expressed in the adult left atrium, and reducing pitx2c expression promotes atrial fibrillation inducibility and complex changes in gene expression. *Circ Cardiovasc Genet*, 4(2):123–33, Apr 2011.

[64] C. A. Klattenhoff, J. C. Scheuermann, L. E. Surface, R. K. Bradley, P. A. Fields, M. L. Steinhauser, H. Ding, V. L. Butty, L. Torrey, S. Haas, R. Abo, M. Tabebordbar, R. T. Lee, C. B. Burge, and L. A. Boyer. Braveheart, a long noncoding rna required for cardiovascular lineage commitment. *Cell*, 152(3):570–83, Jan 2013.

[65] S. W. Kong, Y. W. Hu, J. W. K. Ho, S. Ikeda, S. Polster, R. John, J. L. Hall, E. Bisping, B. Pieske, C. G. dos Remedios, and W. T. Pu. Heart failure-associated changes in rna splicing of sarcomere genes. *Circ Cardiovasc Genet*, 3(2):138–46, Apr 2010.

[66] Kshitiz, M. E. Hubbi, E. H. Ahn, J. Downey, J. Afzal, D.-H. Kim, S. Rey, C. Chang, A. Kundu, G. L. Semenza, R. M. Abraham, and A. Levchenko. Matrix rigidity controls endothelial differentiation and morphogenesis of cardiac precursors. *Sci Signal*, 5(227):ra41, Jun 2012.

[67] J. H. Lee, E. R. Daugharthy, J. Scheiman, R. Kalhor, J. L. Yang, T. C. Ferrante, R. Terry, S. S. F. Jeanty, C. Li, R. Amamoto, D. T. Peters, B. M. Turczyk, A. H. Marblestone, S. A. Inverso, A. Bernard, P. Mali, X. Rios, J. Aach, and G. M. Church. Highly multiplexed subcellular rna sequencing in situ. *Science*, 343(6177):1360–3, Mar 2014.

[68] J.-H. Lee, C. Gao, G. Peng, C. Greer, S. Ren, Y. Wang, and X. Xiao. Analysis of transcriptome complexity through rna sequencing in normal and failing murine hearts. *Circ Res*, 109(12):1332–41, Dec 2011.

[69] J. M. S. Lee and J. M. Kalman. Sinus node dysfunction and atrial fibrillation: two sides of the same coin? *Europace*, 15(2):161–2, Feb 2013.

[70] D. Leung, I. Jung, N. Rajagopal, A. Schmitt, S. Selvaraj, A. Y. Lee, C.-A. Yen, S. Lin, Y. Lin, Y. Qiu, W. Xie, F. Yue, M. Hariharan, P. Ray, S. Kuan, L. Edsall, H. Yang, N. C. Chi, M. Q. Zhang, J. R. Ecker, and B. Ren. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature*, 518(7539):350–4, Feb 2015.

[71] A. L'Honoré, V. Coulon, A. Marcil, M. Lebel, J. Lafrance-Vanasse, P. Gage, S. Camper, and J. Drouin. Sequential expression and redundancy of pitx2 and pitx3 genes during muscle development. *Dev Biol*, 307(2):421–33, Jul 2007.

[72] J. Li, S. Goossens, J. van Hengel, E. Gao, L. Cheng, K. Tyberghein, X. Shang, R. De Rycke, F. van Roy, and G. L. Radice. Loss of αt-catenin alters the hybrid adhering junctions in the heart and leads to dilated cardiomyopathy and ventric-

ular arrhythmia following acute ischemia. *J Cell Sci*, 125(Pt 4):1058–67, Feb 2012.

[73] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–93, Oct 2009.

[74] W.-S. Lin, C.-T. Tai, M.-H. Hsieh, C.-F. Tsai, Y.-K. Lin, H.-M. Tsao, J.-L. Huang, W.-C. Yu, S.-P. Yang, Y.-A. Ding, M.-S. Chang, and S.-A. Chen. Catheter ablation of paroxysmal atrial fibrillation initiated by non-pulmonary vein ectopy. *Circulation*, 107(25):3176–83, Jul 2003.

[75] Z. Lin, A. von Gise, P. Zhou, F. Gu, Q. Ma, J. Jiang, A. L. Yau, J. N. Buck, K. A. Gouin, P. R. R. van Gorp, B. Zhou, J. Chen, J. G. Seidman, D.-Z. Wang, and W. T. Pu. Cardiac-specific yap activation improves cardiac function and survival in an experimental murine mi model. *Circ Res*, 115(3):354–63, Jul 2014.

[76] M. L. Lindsey, D. K. Goshorn, S. Comte-Walters, J. W. Hendrick, E. Hapke, M. R. Zile, and K. Schey. A multidimensional proteomic approach to identify hypertrophy-associated proteins. *Proteomics*, 6(7):2225–35, Apr 2006.

[77] C. Liu, W. Liu, J. Palie, M. F. Lu, N. A. Brown, and J. F. Martin. Pitx2c patterns anterior myocardium and aortic arch vessels and is required for local cell movement into atrioventricular cushions. *Development*, 129(21):5081–91, Nov 2002.

[78] F. Liu, M. D. Levin, N. B. Petrenko, M. M. Lu, T. Wang, L. J. Yuan, A. L. Stout, J. A. Epstein, and V. V. Patel. Histone-deacetylase inhibition reverses atrial arrhythmia inducibility and fibrosis in cardiac hypertrophy independent of angiotensin. *J Mol Cell Cardiol*, 45(6):715–23, Dec 2008.

[79] W. Liu, Y. Komiya, C. Mezzacappa, D. K. Khadka, L. Runnels, and R. Habas. Mim regulates vertebrate neural tube closure. *Development*, 138(10):2035–47, May 2011.

[80] Y. Liu, M. Morley, J. Brandimarto, S. Hannenhalli, Y. Hu, E. A. Ashley, W. H. W. Tang, C. S. Moravec, K. B. Margulies, T. P. Cappola, M. Li, and MAGNet consortium. Rna-seq identifies novel myocardial gene expression signatures of heart failure. *Genomics*, 105(2):83–9, Feb 2015.

[81] Y. Liu-Chittenden, B. Huang, J. S. Shim, Q. Chen, S.-J. Lee, R. A. Anders, J. O. Liu, and D. Pan. Genetic and pharmacological disruption of the tead-yap complex suppresses the oncogenic activity of yap. *Genes Dev*, 26(12):1300–5, Jun 2012.

[82] G. D. Lopaschuk and J. S. Jaswal. Energy metabolic phenotype of the cardiomyocyte during development, differentiation, and postnatal maturation. *J Cardiovasc Pharmacol*, 56(2):130–40, Aug 2010.

[83] S. A. Lubitz, M. F. Sinner, K. L. Lunetta, S. Makino, A. Pfeufer, R. Rahman, C. E. Veltman, J. Barnard, J. C. Bis, S. P. Danik, A. Sonni, M. A. Shea, F. Del Monte, S. Perz, M. Müller, A. Peters, S. M. Greenberg, K. L. Furie, C. van Noord, E. Boerwinkle, B. H. C. Stricker, J. Witteman, J. D. Smith, M. K. Chung, S. R. Heckbert, E. J. Benjamin, J. Rosand, D. E. Arking, A. Alonso, S. Kääb, and P. T. Ellinor. Independent susceptibility markers for atrial fibrillation on chromosome 4q25. *Circulation*, 122(10):976–84, Sep 2010.

[84] S. Mahida and P. T. Ellinor. New advances in the genetic basis of atrial fibrillation. *J Cardiovasc Electrophysiol*, 23(12):1400–6, Dec 2012.

[85] A. Margariti, B. Winkler, E. Karamariti, A. Zampetaki, T.-n. Tsai, D. Baban, J. Ragoussis, Y. Huang, J.-D. J. Han, L. Zeng, Y. Hu, and Q. Xu. Direct reprogramming of fibroblasts into endothelial cells capable of angiogenesis and reendothelialization in tissue-engineered vessels. *Proc Natl Acad Sci U S A*, 109(34):13793–8, Aug 2012.

[86] E. M. McNally, J. R. Golbus, and M. J. Puckelwartz. Genetic mutations and mechanisms in dilated cardiomyopathy. *J Clin Invest*, 123(1):19–26, Jan 2013.

[87] M. R. Mejillano, S.-i. Kojima, D. A. Applewhite, F. B. Gertler, T. M. Svitkina, and G. G. Borisy. Lamellipodial versus filopodial mode of the actin nanomachinery: pivotal role of the filament barbed end. *Cell*, 118(3):363–73, Aug 2004.

[88] T. R. Mercer, S. L. Edwards, M. B. Clark, S. J. Neph, H. Wang, A. B. Stergachis, S. John, R. Sandstrom, G. Li, K. S. Sandhu, Y. Ruan, L. K. Nielsen, J. S. Mattick, and J. A. Stamatoyannopoulos. Dnase i-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat Genet*, 45(8):852–9, Aug 2013.

[89] D. Mosqueira, S. Pagliari, K. Uto, M. Ebara, S. Romanazzo, C. Escobedo-Lucea, J. Nakanishi, A. Taniguchi, O. Franzese, P. Di Nardo, M. J. Goumans, E. Traversa, P. Pinto-do Ó, T. Aoyagi, and G. Forte. Hippo pathway effectors control cardiac progenitor cell fate by acting as dynamic sensors of substrate mechanics and nanostructure. *ACS Nano*, 8(3):2033–47, Mar 2014.

[90] S. L. Murphy, J. Xu, and K. D. Kochanek. Deaths: final data for 2010. *Natl Vital Stat Rep*, 61(4):1–117, May 2013.

[91] C. E. Murry and G. Keller. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell*, 132(4):661–80, Feb 2008.

[92] Y. Nakashima, D. A. Yanez, M. Touma, H. Nakano, A. Jaroszewicz, M. C. Jordan, M. Pellegrini, K. P. Roos, and A. Nakano. Nkx2-5 suppresses the proliferation of atrial myocytes and conduction system. *Circ Res*, 114(7):1103–13, Mar 2014.

[93] S. Neph, J. Vierstra, A. B. Stergachis, A. P. Reynolds, E. Haugen, B. Vernot, R. E. Thurman, S. John, R. Sandstrom, A. K. Johnson, M. T. Maurano, R. Humbert, E. Rynes, H. Wang, S. Vong, K. Lee, D. Bates, M. Diegel, V. Roach, D. Dunn, J. Neri, A. Schafer, R. S. Hansen, T. Kutyavin, E. Giste, M. Weaver, T. Canfield, P. Sabo, M. Zhang, G. Balasundaram, R. Byron, M. J. MacCoss, J. M. Akey, M. A. Bender, M. Groudine, R. Kaul, and J. A. Stamatoyannopoulos. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, 489(7414):83–90, Sep 2012.

[94] C. Newton-Cheh, M. Eijgelsheim, K. M. Rice, P. I. W. de Bakker, X. Yin, K. Estrada, J. C. Bis, K. Marciante, F. Rivadeneira, P. A. Noseworthy, N. Sotoodehnia, N. L. Smith, J. I. Rotter, J. A. Kors, J. C. M. Witteman, A. Hofman, S. R. Heckbert, C. J. O'Donnell, A. G. Uitterlinden, B. M. Psaty, T. Lumley, M. G. Larson, and B. H. C. Stricker. Common variants at ten loci influence qt interval duration in the qtgen study. *Nat Genet*, 41(4):399–406, Apr 2009.

[95] E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, and E. Heard. Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, 485(7398):381–5, May 2012.

[96] A. S. Nord, M. J. Blow, C. Attanasio, J. A. Akiyama, A. Holt, R. Hosseini, S. Phouanenavong, I. Plajzer-Frick, M. Shoukry, V. Afzal, J. L. R. Rubenstein, E. M. Rubin, L. A. Pennacchio, and A. Visel. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell*, 155(7):1521–31, Dec 2013.

[97] C. C. O'Meara, J. A. Wamstad, R. A. Gladstone, G. M. Fomovsky, V. L. Butty, A. Shrikumar, J. B. Gannon, L. A. Boyer, and R. T. Lee. Transcriptional reversion of cardiac myocyte fate during mammalian cardiac regeneration. *Circ Res*, 116(5):804–15, Feb 2015.

[98] S. Ounzain, R. Micheletti, T. Beckmann, B. Schroen, M. Alexanian, I. Pezzuto, S. Crippa, M. Nemir, A. Sarre, R. Johnson, J. Dauvillier, F. Burdet, M. Ibberson, R. Guigó, I. Xenarios, S. Heymans, and T. Pedrazzini. Genome-wide profiling of the cardiac transcriptome after myocardial infarction identifies novel heart-specific long non-coding rnas. *Eur Heart J*, 36(6):353–68a, Feb 2015.

[99] A. Pfeufer, S. Sanna, D. E. Arking, M. Müller, V. Gateva, C. Fuchsberger, G. B. Ehret, M. Orrú, C. Pattaro, A. Köttgen, S. Perz, G. Usala, M. Barbalic, M. Li, B. Pütz, A. Scuteri, R. J. Prineas, M. F. Sinner, C. Gieger, S. S. Najjar, W. H. L. Kao, T. W. Mühleisen, M. Dei, C. Happle, S. Möhlenkamp, L. Crisponi, R. Erbel, K.-H. Jöckel, S. Naitza, G. Steinbeck, F. Marroni, A. A. Hicks, E. Lakatta, B. Müller-Myhsok, P. P. Pramstaller, H.-E. Wichmann, D. Schlessinger, E. Boerwinkle, T. Meitinger, M. Uda, J. Coresh, S. Kääb, G. R. Abecasis, and A. Chakravarti. Common variants at ten loci modulate the qt interval duration in the qtscd study. *Nat Genet*, 41(4):407–14, Apr 2009.

[100] A. Pfeufer, C. van Noord, K. D. Marciante, D. E. Arking, M. G. Larson, A. V. Smith, K. V. Tarasov, M. Müller, N. Sotoodehnia, M. F. Sinner, G. C. Verwo-

ert, M. Li, W. H. L. Kao, A. Köttgen, J. Coresh, J. C. Bis, B. M. Psaty, K. Rice, J. I. Rotter, F. Rivadeneira, A. Hofman, J. A. Kors, B. H. C. Stricker, A. G. Uitterlinden, C. M. van Duijn, B. M. Beckmann, W. Sauter, C. Gieger, S. A. Lubitz, C. Newton-Cheh, T. J. Wang, J. W. Magnani, R. B. Schnabel, M. K. Chung, J. Barnard, J. D. Smith, D. R. Van Wagoner, R. S. Vasan, T. Aspelund, G. Eiriksdottir, T. B. Harris, L. J. Launer, S. S. Najjar, E. Lakatta, D. Schlessinger, M. Uda, G. R. Abecasis, B. Müller-Myhsok, G. B. Ehret, E. Boerwinkle, A. Chakravarti, E. Z. Soliman, K. L. Lunetta, S. Perz, H.-E. Wichmann, T. Meitinger, D. Levy, V. Gudnason, P. T. Ellinor, S. Sanna, S. Kääb, J. C. M. Witteman, A. Alonso, E. J. Benjamin, and S. R. Heckbert. Genome-wide association study of pr interval. *Nat Genet*, 42(2):153–9, Feb 2010.

[101] K. S. Pollard and M. J. van der Laan. Supervised distance matrices. *Stat Appl Genet Mol Biol*, 7(1):Article 33, 2008.

[102] E. R. Porrello, A. I. Mahmoud, E. Simpson, J. A. Hill, J. A. Richardson, E. N. Olson, and H. A. Sadek. Transient regenerative potential of the neonatal mouse heart. *Science*, 331(6020):1078–80, Feb 2011.

[103] B. N. Puente, W. Kimura, S. A. Muralidhar, J. Moon, J. F. Amatruda, K. L. Phelps, D. Grinsfelder, B. A. Rothermel, R. Chen, J. A. Garcia, C. X. Santos, S. Thet, E. Mori, M. T. Kinter, P. M. Rindler, S. Zacchigna, S. Mukherjee, D. J. Chen, A. I. Mahmoud, M. Giacca, P. S. Rabinovitch, A. Aroumougame, A. M. Shah, L. I. Szweda, and H. A. Sadek. The oxygen-rich postnatal environment induces cardiomyocyte cell-cycle arrest through dna damage response. *Cell*, 157(3):565–79, Apr 2014.

[104] Y. Qi, J. A. Ranish, X. Zhu, A. Krones, J. Zhang, R. Aebersold, D. W. Rose, M. G. Rosenfeld, and C. Carrière. Atbf1 is required for the pit1 gene early

activation. *Proc Natl Acad Sci U S A*, 105(7):2481–6, Feb 2008.

[105] L. Qian, E. C. Berry, J.-d. Fu, M. Ieda, and D. Srivastava. Reprogramming of mouse fibroblasts into cardiomyocyte-like cells in vitro. *Nat Protoc*, 8(6):1204–15, Jun 2013.

[106] L. Qian, Y. Huang, C. I. Spencer, A. Foley, V. Vedantham, L. Liu, S. J. Conway, J.-d. Fu, and D. Srivastava. In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. *Nature*, 485(7400):593–8, May 2012.

[107] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang. Genome engineering using the crispr-cas9 system. *Nat Protoc*, 8(11):2281–308, Nov 2013.

[108] S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–80, Dec 2014.

[109] A. J. Ridley. Life at the leading edge. *Cell*, 145(7):1012–22, Jun 2011.

[110] Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari,

K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, and M. Kellis. Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539):317–30, Feb 2015.

[111] K. C. Roberts-Thomson, P. Sanders, and J. M. Kalman. Sinus node disease: an idiopathic right atrial myopathy. *Trends Cardiovasc Med*, 17(6):211–4, Aug 2007.

[112] N. Rosenthal and R. P. Harvey. *Heart development and regeneration*, volume 1. Academic Press, 2010.

[113] J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth, and J. Bustillo. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356):348–52, Jul 2011.

[114] J. Saarikangas, P. K. Mattila, M. Varjosalo, M. Bovellan, J. Hakanen, J. Calzada-Wack, M. Tost, L. Jennen, B. Rathkolb, W. Hans, M. Horsch, M. E. Hyvönen, N. Perälä, H. Fuchs, V. Gailus-Durner, I. Esposito, E. Wolf, M. H. de Angelis,

M. J. Frilander, H. Savilahti, H. Sariola, K. Sainio, S. Lehtonen, J. Taipale, M. Salminen, and P. Lappalainen. Missing-in-metastasis mim/mtss1 promotes actin assembly at intercellular junctions and is required for integrity of kidney epithelia. *J Cell Sci*, 124(Pt 8):1245–55, Apr 2011.

[115] S. Sadasivam, S. Duan, and J. A. DeCaprio. The muvb complex sequentially recruits b-myb and foxm1 to promote mitotic gene expression. *Genes Dev*, 26(5):474–89, Mar 2012.

[116] K. J. Schillinger and V. V. Patel. Atrial fibrillation in the elderly: the potential contribution of reactive oxygen species. *J Geriatr Cardiol*, 9(4):379–88, Dec 2012.

[117] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli. Three-dimensional folding and functional organization principles of the drosophila genome. *Cell*, 148(3):458–72, Feb 2012.

[118] T. Shen, I. Aneas, N. Sakabe, R. J. Dirschinger, G. Wang, S. Smemo, J. M. Westlund, H. Cheng, N. Dalton, Y. Gu, C. J. Boogerd, C.-l. Cai, K. Peterson, J. Chen, M. A. Nobrega, and S. M. Evans. Tbx20 regulates a genetic program essential to adult mouse cardiomyocyte function. *J Clin Invest*, 121(12):4640–54, Dec 2011.

[119] Y. Shen, F. Yue, D. F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenkov, and B. Ren. A map of the cis-regulatory sequences in the mouse genome. *Nature*, 488(7409):116–20, Aug 2012.

[120] S. Smemo, J. J. Tena, K.-H. Kim, E. R. Gamazon, N. J. Sakabe, C. Gómez-Marín, I. Aneas, F. L. Credidio, D. R. Sobreira, N. F. Wasserman, J. H. Lee, V. Puviindran, D. Tam, M. Shen, J. E. Son, N. A. Vakili, H.-K. Sung, S. Naranjo,

R. D. Acemel, M. Manzanares, A. Nagy, N. J. Cox, C.-C. Hui, J. L. Gomez-Skarmeta, and M. A. Nóbrega. Obesity-associated variants within fto form long-range functional connections with irx3. *Nature*, 507(7492):371–5, Mar 2014.

[121] G. K. Smyth. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, 3:Article3, 2004.

[122] S. Sood, M. G. Chelu, R. J. van Oort, D. Skapura, M. Santonastasi, D. Dobrev, and X. H. T. Wehrens. Intracellular calcium leak due to fkbp12.6 deficiency in mice facilitates the inducibility of atrial fibrillation. *Heart Rhythm*, 5(7):1047–54, Jul 2008.

[123] N. Sotoodehnia, A. Isaacs, P. I. W. de Bakker, M. Dörr, C. Newton-Cheh, I. M. Nolte, P. van der Harst, M. Müller, M. Eijgelsheim, A. Alonso, A. A. Hicks, S. Padmanabhan, C. Hayward, A. V. Smith, O. Polasek, S. Giovannone, J. Fu, J. W. Magnani, K. D. Marciante, A. Pfeufer, S. A. Gharib, A. Teumer, M. Li, J. C. Bis, F. Rivadeneira, T. Aspelund, A. Köttgen, T. Johnson, K. Rice, M. P. S. Sie, Y. A. Wang, N. Klopp, C. Fuchsberger, S. H. Wild, I. Mateo Leach, K. Estrada, U. Völker, A. F. Wright, F. W. Asselbergs, J. Qu, A. Chakravarti, M. F. Sinner, J. A. Kors, A. Petersmann, T. B. Harris, E. Z. Soliman, P. B. Munroe, B. M. Psaty, B. A. Oostra, L. A. Cupples, S. Perz, R. A. de Boer, A. G. Uitterlinden, H. Völzke, T. D. Spector, F.-Y. Liu, E. Boerwinkle, A. F. Dominiczak, J. I. Rotter, G. van Herpen, D. Levy, H.-E. Wichmann, W. H. van Gilst, J. C. M. Witteman, H. K. Kroemer, W. H. L. Kao, S. R. Heckbert, T. Meitinger, A. Hofman, H. Campbell, A. R. Folsom, D. J. van Veldhuisen, C. Schwienbacher, C. J. O'Donnell, C. B. Volpato, M. J. Caulfield, J. M. Connell, L. Launer, X. Lu, L. Franke, R. S. N. Fehrmann, G. te Meer-

man, H. J. M. Groen, R. K. Weersma, L. H. van den Berg, C. Wijmenga, R. A. Ophoff, G. Navis, I. Rudan, H. Snieder, J. F. Wilson, P. P. Pramstaller, D. S. Siscovick, T. J. Wang, V. Gudnason, C. M. van Duijn, S. B. Felix, G. I. Fishman, Y. Jamshidi, B. H. C. Stricker, N. J. Samani, S. Kääb, and D. E. Arking. Common variants in 22 loci are associated with qrs duration and cardiac ventricular conduction. *Nat Genet*, 42(12):1068–76, Dec 2010.

[124] E. Splinter, E. de Wit, H. J. G. van de Werken, P. Klous, and W. de Laat. Determining long-range chromatin interactions for selected genomic sites using 4c-seq technology: from fixation to computation. *Methods*, 58(3):221–30, Nov 2012.

[125] R. Stadhouders, P. Kolovos, R. Brouwer, J. Zuin, A. van den Heuvel, C. Kockx, R.-J. Palstra, K. S. Wendt, F. Grosveld, W. van Ijcken, and E. Soler. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc*, 8(3):509–24, Mar 2013.

[126] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131(5):861–72, Nov 2007.

[127] K. Takahashi and S. Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–76, Aug 2006.

[128] S. Tanigawa, C. H. Lee, C. S. Lin, C. C. Ku, H. Hasegawa, S. Qin, A. Kawahara, Y. Korenori, K. Miyamori, M. Noguchi, L. H. Lee, Y. C. Lin, C. L. Steve Lin, Y. Nakamura, C. Jin, N. Yamaguchi, R. Eckner, D.-X. Hou, and K. K. Yokoyama. Jun dimerization protein 2 is a critical component of the

nrf2/mafk complex regulating the response to ros homeostasis. *Cell Death Dis*, 4:e921, 2013.

[129] E. Tarazón, E. Roselló-Lletí, M. Rivera, A. Ortega, M. M. Molina-Navarro, J. C. Triviño, F. Lago, J. R. González-Juanatey, P. Orosa, J. A. Montero, A. Salvador, and M. Portolés. Rna sequencing analysis and atrial natriuretic peptide production in patients with dilated and ischemic cardiomyopathy. *PLoS One*, 9(3):e90157, 2014.

[130] R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutyavin, B. Lajoie, B.-K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, and J. A. Stamatoyannopoulos. The accessible chromatin landscape of the human genome. *Nature*, 489(7414):75–82, Sep 2012.

[131] K. Tschöp, A. R. Conery, L. Litovchick, J. A. Decaprio, J. Settleman, E. Harlow, and N. Dyson. A kinase shrna screen links lats2 and the prb tumor suppressor. *Genes Dev*, 25(8):814–30, Apr 2011.

[132] J. H. van Berlo, M. Maillet, and J. D. Molkentin. Signaling effectors underlying pathologic growth and remodeling of the heart. *J Clin Invest*, 123(1):37–45, Jan 2013.

[133] V. Vedantham, G. Galang, M. Evangelista, R. C. Deo, and D. Srivastava. Rna sequencing of mouse sinoatrial node reveals an upstream regulatory role for islet-1 in cardiac pacemaker cells. *Circ Res*, 116(5):797–803, Feb 2015.

[134] A. von Gise, Z. Lin, K. Schlegelmilch, L. B. Honor, G. M. Pan, J. N. Buck, Q. Ma, T. Ishiwata, B. Zhou, F. D. Camargo, and W. T. Pu. Yap1, the nuclear target of hippo signaling, stimulates heart growth through cardiomyocyte proliferation but not hypertrophy. *Proc Natl Acad Sci U S A*, 109(7):2394–9, Feb 2012.

[135] G. Q. Wallace and E. M. McNally. Mechanisms of muscle degeneration, regeneration, and repair in the muscular dystrophies. *Annu Rev Physiol*, 71:37–57, 2009.

[136] J. A. Wamstad, J. M. Alexander, R. M. Truty, A. Shrikumar, F. Li, K. E. Eilertson, H. Ding, J. N. Wylie, A. R. Pico, J. A. Capra, G. Erwin, S. J. Kattman, G. M. Keller, D. Srivastava, S. S. Levine, K. S. Pollard, A. K. Holloway, L. A. Boyer, and B. G. Bruneau. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell*, 151(1):206–20, Sep 2012.

[137] J. A. Wamstad, X. Wang, O. O. Demuren, and L. A. Boyer. Distal enhancers: new insights into heart development and disease. *Trends Cell Biol*, 24(5):294–302, May 2014.

[138] J. Wang, E. Klysik, S. Sood, R. L. Johnson, X. H. T. Wehrens, and J. F. Martin. Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proc Natl Acad Sci U S A*, 107(21):9753–8, May 2010.

[139] J. Wang and J. F. Martin. Macro advances in micrornas and myocardial regeneration. *Curr Opin Cardiol*, 29(3):207–13, May 2014.

[140] S. Wang, A. L. Lopez, 3rd, Y. Morikawa, G. Tao, J. Li, I. V. Larina, J. F. Martin, and K. V. Larin. Noncontact quantitative biomechanical characterization of cardiac muscle using shear wave imaging optical coherence tomography. *Biomed Opt Express*, 5(7):1980–92, Jul 2014.

[141] Z.-B. Wang, Z.-Z. Jiang, Q.-H. Zhang, M.-W. Hu, L. Huang, X.-H. Ou, L. Guo, Y.-C. Ouyang, Y. Hou, C. Brakebusch, H. Schatten, and Q.-Y. Sun. Specific deletion of cdc42 does not affect meiotic spindle organization/migration and homologous chromosome segregation but disrupts polarity establishment and cytokinesis in mouse oocytes. *Mol Biol Cell*, 24(24):3832–41, Dec 2013.

[142] M. T. Weirauch, A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.-C. Lozano, M. Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F.-Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker, and T. R. Hughes. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*, 158(6):1431–43, Sep 2014.

[143] M. A. Wheeler, A. Warley, R. G. Roberts, E. Ehler, and J. A. Ellis. Identification of an emerin-beta-catenin complex in the heart important for intercalated disc architecture and beta-catenin localisation. *Cell Mol Life Sci*, 67(5):781–96, Mar 2010.

[144] A. A. M. Wilde and R. Brugada. Phenotypical manifestations of mutations in the genes encoding subunits of the cardiac sodium channel. *Circ Res*, 108(7):884–97, Apr 2011.

[145] H. Wu, A. S. Nord, J. A. Akiyama, M. Shoukry, V. Afzal, E. M. Rubin, L. A. Pennacchio, and A. Visel. Tissue-specific rna expression marks distant-acting

developmental enhancers. *PLoS Genet*, 10(9):e1004610, Sep 2014.

[146] Z. Xia, L. A. Donehower, T. A. Cooper, J. R. Neilson, D. A. Wheeler, E. J. Wagner, and W. Li. Dynamic analyses of alternative polyadenylation from rna-seq reveal a 3'-utr landscape across seven tumour types. *Nat Commun*, 5:5274, 2014.

[147] Y. Xiao, M. Zhang, Y. Morikawa, and J. F. Martin. Hippo and mechanical signaling cooperatively determine epicardial progenitor cell fate. *Submitted*, 2015.

[148] M. Xin, Y. Kim, L. B. Sutherland, M. Murakami, X. Qi, J. McAnally, E. R. Porrello, A. I. Mahmoud, W. Tan, J. M. Shelton, J. A. Richardson, H. A. Sadek, R. Bassel-Duby, and E. N. Olson. Hippo pathway effector yap promotes cardiac regeneration. *Proc Natl Acad Sci U S A*, 110(34):13839–44, Aug 2013.

[149] C. Xing, H. Xie, L. Zhou, W. Zhou, W. Zhang, S. Ding, B. Wei, X. Yu, R. Su, and S. Zheng. Cyclin-dependent kinase inhibitor 3 is overexpressed in hepato-cellular carcinoma and promotes tumor cell proliferation. *Biochem Biophys Res Commun*, 420(1):29–35, Mar 2012.

[150] F.-X. Yu and K.-L. Guan. The hippo pathway: regulators and regulations. *Genes Dev*, 27(4):355–71, Feb 2013.

[151] F. Yue, Y. Cheng, A. Breschi, J. Vierstra, W. Wu, T. Ryba, R. Sandstrom, Z. Ma, C. Davis, B. D. Pope, Y. Shen, D. D. Pervouchine, S. Djebali, R. E. Thurman, R. Kaul, E. Rynes, A. Kirilusha, G. K. Marinov, B. A. Williams, D. Trout, H. Amrhein, K. Fisher-Aylor, I. Antoshechkin, G. DeSalvo, L.-H. See, M. Fastuca, J. Drenkow, C. Zaleski, A. Dobin, P. Prieto, J. Lagarde, G. Bus-sotti, A. Tanzer, O. Denas, K. Li, M. A. Bender, M. Zhang, R. Byron, M. T. Groudine, D. McCleary, L. Pham, Z. Ye, S. Kuan, L. Edsall, Y.-C. Wu, M. D.

Rasmussen, M. S. Bansal, M. Kellis, C. A. Keller, C. S. Morrissey, T. Mishra, D. Jain, N. Dogan, R. S. Harris, P. Cayting, T. Kawli, A. P. Boyle, G. Euskirchen, A. Kundaje, S. Lin, Y. Lin, C. Jansen, V. S. Malladi, M. S. Cline, D. T. Erickson, V. M. Kirkup, K. Learned, C. A. Sloan, K. R. Rosenbloom, B. Lacerda de Sousa, K. Beal, M. Pignatelli, P. Flicek, J. Lian, T. Kahveci, D. Lee, W. J. Kent, M. Ramalho Santos, J. Herrero, C. Notredame, A. Johnson, S. Vong, K. Lee, D. Bates, F. Neri, M. Diegel, T. Canfield, P. J. Sabo, M. S. Wilken, T. A. Reh, E. Giste, A. Shafer, T. Kutyavin, E. Haugen, D. Dunn, A. P. Reynolds, S. Neph, R. Humbert, R. S. Hansen, M. De Bruijn, L. Selleri, A. Rudensky, S. Josefowicz, R. Samstein, E. E. Eichler, S. H. Orkin, D. Levasseur, T. Papayannopoulou, K.-H. Chang, A. Skoultchi, S. Gosh, C. Disteche, P. Treuting, Y. Wang, M. J. Weiss, G. A. Blobel, X. Cao, S. Zhong, T. Wang, P. J. Good, R. F. Lowdon, L. B. Adams, X.-Q. Zhou, M. J. Pazin, E. A. Feingold, B. Wold, J. Taylor, A. Mortazavi, S. M. Weissman, J. A. Stamatoyannopoulos, M. P. Snyder, R. Guigo, T. R. Gingeras, D. M. Gilbert, R. C. Hardison, M. A. Beer, B. Ren, and Mouse ENCODE Consortium. A comparative encyclopedia of dna elements in the mouse genome. *Nature*, 515(7527):355–64, Nov 2014.

[152] A. C. Zambon, S. Gaj, I. Ho, K. Hanspers, K. Vranizan, C. T. Evelo, B. R. Conklin, A. R. Pico, and N. Salomonis. Go-elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics*, 28(16):2209–10, Aug 2012.

[153] A. Zemljic-Harpf, A. M. Manso, and R. S. Ross. Vinculin and talin: focus on the myocardium. *J Investig Med*, 57(8):849–55, Dec 2009.

[154] Z. Zhang, Y. He, D. Tuteja, D. Xu, V. Timofeyev, Q. Zhang, K. A. Glatter, Y. Xu, H.-S. Shin, R. Low, and N. Chiamvimonvat. Functional roles of cav1.3(alpha1d) calcium channels in atria: insights gained from gene-targeted

null mutant mice. *Circulation*, 112(13):1936–44, Sep 2005.

[155] P. Zhou, Y. Zhang, Q. Ma, F. Gu, D. S. Day, A. He, B. Zhou, J. Li, S. M. Stevens, D. Romo, and W. T. Pu. Interrogating translational efficiency and lineage-specific transcriptomes using ribosome affinity purification. *Proc Natl Acad Sci U S A*, 110(38):15395–400, Sep 2013.

[156] Y. Zhou, G. Cheng, L. Dieter, T. A. Hjalt, F. H. Andrade, J. S. Stahl, and H. J. Kaminski. An altered phenotype in a conditional knockout of pitx2 in extraocular muscle. *Invest Ophthalmol Vis Sci*, 50(10):4531–41, Oct 2009.

[157] Y. Zhou, B. Gong, and H. J. Kaminski. Genomic profiling reveals pitx2 controls expression of mature extraocular muscle contraction-related genes. *Invest Ophthalmol Vis Sci*, 53(4):1821–9, Apr 2012.